Instanton Counting, Matrix Models, and Characters

Spencer Tamagni
University of Central Florida

INSTANTON COUNTING, MATRIX MODELS, AND CHARACTERS

by

SPENCER TAMAGNI

A thesis submitted in partial fulfilment of the requirements
for the Honors in the Major Program in Physics
in the College of Sciences
and in the Burnett Honors College
at the University of Central Florida

Spring Term
2022

Thesis Chair: Costas Efthimiou

ABSTRACT

In this thesis we study symmetries of quantum field theory visible only at the non-perturbative level, which arise from certain large deformations of the path integration contour. We exposit the recently-developed theory of qq-characters [40] that organizes such symmetries in the case of $\mathscr{N} = 2$ supersymmetric gauge theories in four dimensions. We sketch the physical origin of such observables from intersecting branes in string theory, and the mathematical origin as certain equivariant integrals over Nakajima quiver varieties. We explain the main applications, including the derivation of Seiberg-Witten geometry for quiver gauge theories and the relations to quantum integrable systems.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Quantum field theory is an essential component of the description of nature at the fundamental level, as it is the mathematical language used to describe the interactions of elementary particles. It also is a natural framework for providing low-energy effective descriptions of various systems relevant to condensed matter physics. Despite its ubiquity in theoretical physics, robust exact results in quantum field theory are rare and signify that there is still much to be learned in this mature discipline.

Because of this lack of exact results, it is desirable to search for simple examples of field theories in which exact calculations can be performed. During the past twenty years of research in mathematical physics, much progress has been made on developing effective methods of exact calculation in $\mathscr{N} = 2$ supersymmetric gauge theories in four dimensions (see e.g. [54] for a review of recent developments). We emphasize that in our study, the assumption of supersymmetry is a theoretical idealization used to formulate tractable models, not a prediction that such a symmetry is realized in nature. The key to exact calculations in these theories is the supersymmetric localization technique (closely related to the possibility of performing the topological twist introduced by Witten [60]). This allows a certain sector of observables, known as the BPS sector, to be computed exactly by reduction of the path integral to a finite-dimensional integral. After subjecting these theories to the so-called Ω-deformation [39], [45], the resulting finite-dimensional integrals can be calculated using the Atiyah-Bott localization formula in equivariant cohomology [1], [3], [13] which reduces the integrals to sums over a set of fixed points of a group action.

Because of the ability to analyze $\mathscr{N} = 2$ theories at an increased level of detail, it is desirable to search for evidence of deeper symmetries present within them. One type of symmetry which would ordinarily be out of reach in quantum field theory is a symmetry related to changing the instanton charge, that is, the topological sector of the field space (in the case of gauge theories in flat space this means changing the second Chern class of the gauge field configuration). One finds that the Ward identities for these symmetries of $\mathscr{N}=2$ theories, referred to as non-perturbative Dyson Schwinger equations, are naturally organized via a structure known as the *qq*-characters [40]. Strictly speaking, this structure is understood for only a subclass of all possible $\mathscr{N}=2$ theories known as quiver gauge theories. In this thesis, we will consider only quiver gauge theories.

The *qq*-characters are remarkable objects with diverse applications. In this thesis, we exposit their origin in the study of gauge theories. We begin in chapter 2 with a mathematical motivation, explaining some geometric aspects of representation theory. In chapter 3 we review aspects of the solution of random matrix models at large N, as many of the gauge theory constructions can be motivated as an effort to reproduce this kind of analysis in the case of $\mathscr{N}=2$ gauge theories. In chapter 4 we recall the basics of $\mathscr{N}=2$ supersymmetric gauge theories and define the quiver gauge theories under study. In chapter 5 we explain the Atiyah-Drinfeld-Hitchin-Manin (ADHM) construction of instantons [2] and the computation of the instanton partition function of the $\mathscr{N}=2$ theories in the Ω-background, the so-called Nekrasov partition function [39]. Finally, in chapter 6 we explain the generalization of the ADHM construction to the so-called moduli space of crossed instantons [41] and introduce the *qq*-characters [40]. We explain some of the applications, including the Seiberg-Witten solutions of the theories under study and a connection with integrable systems via consideration of surface operators. The main objective of this thesis is the discussion of chapter 6 reviewing *qq*-characters starting from their origin in geometry and gauge theory; the majority of the document is an explanation of the technical background required to follow this discussion.

CHAPTER 2: CHARACTERS AND GEOMETRY

In this chapter we provide mathematical context for some of the structures to appear later within this thesis, namely relations between representation theory (specifically the theory of characters) and geometry. The purpose is to provide mathematically inclined readers with some motivation; for readers interested purely in quantum field theory this chapter may be skipped. As this chapter is purely motivational, we do not attempt to make completely rigorous or general mathematical statements.

Characters from Quantum Mechanics

Representation theory is the study of group actions on vector spaces by linear transformations, say a group G acting on a vector space V (in this thesis we take all vector spaces to be complex). A natural origin for such a situation within physics comes from quantum mechanics, where G may be identified with the group of symmetries of a particular quantum mechanical system and V is identified with the Hilbert space of the system in question. It is often the case that such a quantum mechanical system arises from quantizing a classical phase space (X,ω), where X is a symplectic manifold with symplectic form ω, such that the group G acts on X in a way that preserves ω.

Given a representation V of G, which is realized as a homomorphism $\rho : G \to GL(V)$ where $GL(V)$ denotes the group of invertible linear maps from V to itself, one constructs the character

$$\chi_V(g) = \mathrm{tr}_V\, \rho(g). \tag{2.1}$$

The character itself defines a homomorphism from the representation ring of G to the ring of functions on G which are constant along conjugacy classes. If one denotes a set of local Darboux coordinates for (X,ω) schematically by (q,p), then the phase space path integral expresses the character (at least formally) in terms of the geometry of X:

$$\chi_V(g) = \int_{\gamma(1)=g\cdot\gamma(0)} \mathscr{D}p(t)\mathscr{D}q(t)\exp\left\{ i\int \gamma^*\theta \right\} \tag{2.2}$$

where θ is a locally-defined one-form on X such that $\omega = d\theta$, and the integral is evaluated over the space of maps $\gamma\colon [0,1] \to X$, where $\gamma(t) = (q(t), p(t))$, subject to the twisted boundary condition $\gamma(1) = g\cdot\gamma(0)$. In certain situations this path integral can be reduced to a finite-dimensional integral.

One such situation is when G is a compact simple Lie group, and $X = G/T$ is the so-called complete flag variety of G (T denotes a maximal torus of G). Then X is in fact a Kähler manifold (in fact a projective algebraic variety), and one can apply the geometric quantization prescription [63], [33] to quantize it. The cohomology class of the symplectic form ω is the first Chern class of a G-equivariant holomorphic line bundle $\mathscr{L}$ over X–the integrality condition (Bohr-Sommerfeld condition) in geometric quantization determines the topological isomorphism class of $\mathscr{L}$ uniquely in terms of a weight λ of G, so we denote the line bundle by $\mathscr{L}_\lambda$. The Borel-Weil-Bott construction [52], [4] claims that, so long as λ satisfies a positivity condition, the space of holomorphic sections $H^0(X,\mathscr{L}_\lambda)$ is an irreducible representation of G with highest weight λ. $H^0(X,\mathscr{L}_\lambda)$ is also identified with the Hilbert space in the geometric quantization setup, so in this way one produces irreducible representations of compact Lie groups by quantizing classical phase spaces.

If $\mathscr{L}_\lambda$ is sufficently positive, then by the Kodaira vanishing theorem [22] the higher cohomologies vanish, $H^i(X,\mathscr{L}_\lambda) = 0$ for $i \geq 1$. The character of the representation $V = H^0(X,\mathscr{L}_\lambda)$ can then be evaluated using the equivariant version of the Grothendieck-Riemann-Roch formula. Because characters are constant along conjugacy classes, it is enough to evaluate the character on a generic element $t \in T \subset G$. We abuse notation and identify t with the linear transformation representing its action on a given vector space, in a hopefully obvious sense; then the T-equivariant index formula reads

$$\chi_V(t) = \mathrm{tr}_{H^0(X,\mathscr{L}_\lambda)}\, t = \sum_{i\geq 0} (-1)^i \,\mathrm{tr}_{H^i(X,\mathscr{L}_\lambda)}\, t = \int_X \mathrm{td}_T(X)\mathrm{ch}_T(\mathscr{L}_\lambda). \tag{2.3}$$

td_T and ch_T denote the T-equivariant Todd class and Chern character, respectively, see for example [50]. The finite-dimensional integral on the right hand side of the above equation can be interpreted as a reduction of the phase space path integral using supersymmetric localization[1] but we will not explore this interpretation further.

The objects $\mathrm{td}_T(X)$ and $\mathrm{ch}_T(\mathscr{L}_\lambda)$ define elements of the equivariant cohomology ring $H_T^*(X)$. Then the integral from equation (2.3) may be evaluated by the Atiyah-Bott fixed point formula [1], and reduces to a sum over the fixed points of T acting on X. It is easy to verify that such fixed points are in one-to-one correspondence with the elements of the Weyl group of G, denoted by W, and if $g = \exp(u)$ for $u \in \mathrm{Lie}(G)$ then the weights of the group action in the holomorphic tangent space to X at the fixed point corresponding to $w \in W$ are $(w\cdot\alpha)(u)$, where $\alpha \in \mathrm{Lie}(G)^*$ ranges over the positive roots of $\mathrm{Lie}(G)$ and $w\cdot\alpha$ denotes the action of the element $w \in W$ on α. Then the fixed point formula reads (we use the notation $t^a := \exp(a(u))$ for $a \in \mathrm{Lie}(G)^*$):

$$\chi_V(t) = \sum_{w\in W} \frac{t^{w\cdot\lambda}}{\prod_{\alpha>0}(1-t^{-w\cdot\alpha})} = \sum_{w\in W} \varepsilon(w) \frac{t^{w\cdot(\lambda+\rho)}}{\prod_{\alpha>0}(t^{\alpha/2}-t^{-\alpha/2})}. \tag{2.4}$$

This is simply the Weyl character formula, where we denote the Weyl vector by ρ and $\varepsilon : W \to \mathbb{Z}_2$ is the sign homomorphism. We write $\alpha > 0$ as a shorthand for α being a positive root.

[1]The problem has no obvious supersymmetry, but it enters upon recalling that the phase space path integration measure is defined via the Liouville measure on X which is essentially the Pfaffian of ω; upon representing this Pfaffian via a fermionic path integral this system has a supersymmetry which acts on the fields as the equivariant de Rham differential of the loop space of X, which is equivariant with respect to the $U(1)$ symmetry rotating the loops.

The discussion in the previous section has produced characters of highest weight representations of a compact Lie group G by applying geometric quantization to the complete flag variety G/T. The vector space furnishing the representation was obtained as a sheaf cohomology group $H^0(G/T, \mathscr{L}_\lambda)$, utilizing the complex structure of G/T in an essential way.

In (relatively) recent years, a new method to construct representations of Lie algebras has emerged, which uses topological cohomology groups as opposed to sheaf cohomology groups, and replaces the flag varieties G/T by other objects with richer structure. For the purposes of this motivational section, we do not explain the general theory, rather focusing on the special case of the $\mathfrak{sl}_2$ Lie algebra.

Fix an integer $w \geq 1$, and consider the smooth complex manifold $T^*Gr(v,w)$, for some integer v such that $0 \leq v \leq w$. $Gr(v,w)$ denotes the Grassmannian of v-planes in $\mathbb{C}^w$, and T^* denotes the holomorphic cotangent bundle. This space is in fact a smooth hyper-Kähler manifold. Consider the vector space

$$V = \bigoplus_{0 \leq v \leq w} H^{\mathrm{mid}}(T^*Gr(v,w)) \tag{2.5}$$

where H^{mid} denotes the middle-dimensional topological cohomology. Then it is known (see [36] for a review) that V is an irreducible representation of $\mathfrak{sl}_2$ with spin $w/2$. A key feature of the construction is that the algebra of $\mathfrak{sl}_2$ acts by correspondences, that is to say, subvarieties $Z \subset T^*Gr(v,w) \times T^*Gr(v',w)$ determine linear maps $\delta_Z : H^{\mathrm{mid}}(T^*Gr(v,w)) \to H^{\mathrm{mid}}(T^*Gr(v',w))$ via Poincaré duality and intersection pairing, and suitable choices of Z produce in this way generators of $\mathfrak{sl}_2$ acting on V.

The manifolds $T^*Gr(v,w)$ carry further structure: they are acted on by the group $GL(w) \times \mathbb{C}^\times$, where the $GL(w)$ action is induced by the natural one on $\mathbb{C}^w$ and $\mathbb{C}^\times$ acts by scaling the fibers of the cotangent bundle. One can consider another vector space, obtained by taking equivariant cohomology:

$$\bigoplus_{0\leq v\leq w} H^*_{GL(w)\times\mathbb{C}^\times}(T^*Gr(v,w)). \tag{2.6}$$

This vector space is a finite dimensional representation of an infinite-dimensional algebra known as the Yangian of $\mathfrak{sl}_2$, denoted by $Y(\mathfrak{sl}_2)$ [58]. There is a further generalization where one considers the $GL(w)\times\mathbb{C}^\times$-equivariant K-theory of (the disjoint union over v of) $T^*Gr(v,w)$, which turns out to be a representation of the so-called quantum affine algebra $U_q(\widehat{\mathfrak{sl}}_2)$ [37].

Since the Borel-Weil-Bott method of producing representations has a transparent physical interpretation in terms of a quantization procedure, it is natural to ask how the above kinds of representations can be produced using physical ideas. In this thesis, we will present an example of a physical system where these structures appear, the so-called $\mathcal{N}=2$ supersymmetric gauge theory with gauge group $SU(n)$ coupled to $2n$ hypermultiplets in the fundamental representation. In this construction, the Yangian $Y(\mathfrak{sl}_2)$ appears to be intimately related to deep symmetries of quantum field theory. In fact, the quantum field theory constructions appear to generalize this story by introducing an additional deformation parameter.

Another (more mathematical) question is if there is an analog of the formula (2.3) expressing the character of the Borel-Weil-Bott representation in terms of the geometry of G/T via some kind of equivariant integral (for us, an equivariant integral is an integral of an equivariant cohomology class). One would like to express the character of the representations explained in this section in terms of the geometry of $T^*Gr(v,w)$. Quantum field theory once again produces such a formula.

CHAPTER 3: MATRIX MODELS AND LIMIT SHAPES

Before embarking on the study of the theories of interest, in this chapter we review certain constructions in the study of matrix models, that is, gauge theories in zero dimensions of spacetime. These models are of great interest in their own right, but for our purposes they will serve as an inspiration to study certain constructions in four-dimensional gauge theories.

Planar Diagrams and Spectral Curve

Matrix models are interesting toy models of quantum field theory because they are the simplest theories with nontrivial large-N expansion, which can in turn be solved exactly at large N. The exact solution involves the (classical) algebraic geometry of a Riemann surface known as the spectral curve that characterizes the limiting behavior of the distribution of eigenvalues of these random matrices as the size N of the matrix is taken to infinity. We briefly review this set of constructions here. We provide some detail because a similar structure will be found in the gauge theories of interest.

The partition function of the matrix model is defined as

$$Z = \frac{1}{\mathrm{vol}\, U(N)} \int [d\Phi] e^{-\frac{1}{g_s} \mathrm{tr} V(\Phi)}. \tag{3.1}$$

Here, Φ is an $N \times N$ hermitian matrix and $V(\Phi)$ is an arbitrary polynomial of some degree d, called the potential, which also plays the role of the action in this zero-dimensional theory. The action is clearly invariant under $\Phi \to U\Phi U^\dagger$ for $U \in U(N)$, a zero-dimensional gauge transformation. The parameter g_s is known as the string coupling constant. If this were regarded as a zero-dimensional Yang-Mills theory, the conventional gauge theory coupling constant would be related to it by $g_s =$

g_{YM}^2. It also plays the role of Planck's constant, controlling the semiclassical expansion. The measure is given by

$$[d\Phi] = 2^{\frac{N(N-1)}{2}} \prod_{i=1}^{N} d\Phi_{ii} \prod_{1\leq i<j\leq N} d\,\mathrm{Re}\,\Phi_{ij} d\,\mathrm{Im}\,\Phi_{ij}. \tag{3.2}$$

The volume of the gauge group simply sets the natural normalization for the gauge theory path integral. It is straightforward to show that

$$\mathrm{vol}\,U(N) = \frac{(2\pi)^{\frac{N(N+1)}{2}}}{1!\cdot 2!\cdot \ldots \cdot (N-1)!}. \tag{3.3}$$

The free energy $F = \log Z$ of the matrix model is given as a sum over connected Feynman diagrams, with Feynman rules depending on the choice of potential. In 't Hooft's double-line notation (see e.g. [24], [64], [29] for a review), a Feynman graph with h index loops can be regarded as a topological surface of some genus g with h disks removed, which has Euler characteristic $\chi = 2-2g-h$. Summing over all topological types and denoting by $F_{g,h}$ the amplitude associated with the graphs with genus g and h holes, the free energy has an expansion

$$F(g_s,N) = \sum_{g\geq 0}\sum_{h\geq 1} g_s^{2g-2+h} F_{g,h} N^h. \tag{3.4}$$

Following 't Hooft, one considers the limit $N\to\infty$, $g_s\to 0$, with $t = g_s N$ fixed. Then the free energy has an expansion

$$F(g_s,t) = \sum_{g\geq 0} g_s^{2g-2} F_g(t) \tag{3.5}$$

with

$$F_g(t) = \sum_{h\geq 1} F_{g,h} t^h. \tag{3.6}$$

The $1/N$ expansion of the matrix model is naturally expressed as a sum over closed surfaces, with the N^{2-2g} contribution associated to a surface of genus g. In the $N\to\infty$ limit at fixed t, the surfaces of genus zero dominate, and their contribution is given as the sum of all planar Feynman diagrams in the matrix model.

These are statements of general validity in a large class of quantum field theories, and do not provide much insight into solving the matrix model. We turn our attention to this issue. From equation (3.1) it is clear that the large N limit at fixed t, in which $g_s \to 0$, corresponds to evaluating the integral by saddle point approximation. To facilitate this one takes advantage of the $U(N)$ symmetry to convert the integral over matrices Φ to an integral over the eigenvalues λ_i, which introduces a Jacobian factor given by the standard Vandermonde determinant. Equation (3.1) becomes, in these variables,

$$Z = \frac{1}{N!}\int \frac{d^N\lambda}{(2\pi)^N} \prod_{1\leq i<j\leq N} (\lambda_i - \lambda_j)^2 \exp\left\{-\frac{1}{g_s}\sum_{i=1}^{N} V(\lambda_i)\right\}. \tag{3.7}$$

The integrand may be written as $e^{-\frac{N^2}{t}S_{\text{eff}}}$ where we have introduced the effective action or Boltzmann weight

$$S_{\text{eff}}(\lambda_1, ..., \lambda_N) = \frac{1}{N}\sum_{i=1}^{N} V(\lambda_i) - \frac{g_s}{N}\sum_{i<j}\log(\lambda_i - \lambda_j)^2. \tag{3.8}$$

Note that the sum of N terms of order 1 is of order N, and recalling $g_s \propto 1/N$ one sees that the effective action is $\mathscr{O}(1)$ as $N \to \infty$ so the saddle point approximation is applicable to evaluate the integral. The saddle point method is best thought of as taking place in the complex plane, so the eigenvalues should be taken as complex and one should search for all possible complex critical points–this is equivalent to allowing for a deformation of the contour in the integral over λ_i.

Differentiating (3.8) one finds that the saddle points must satisfy

$$V'(\lambda_i) = 2g_s \sum_{j(\neq i)} \frac{1}{\lambda_i - \lambda_j}. \tag{3.9}$$

This equation describes an equilibrium condition for a Coulomb gas of N eigenvalues in the plane subject to a one-body potential $V(x)$. Eigenvalues tend to condense near the critical points of V, spreading out into line segments due to the Coulomb repulsion.

To confirm this naive expectation, introduce the so-called resolvent $W_1(x)$ via the equation (brackets denote the expectation value with respect to the measure defined via the matrix model path integral)

$$g_s W_1(x) = \left\langle g_s \operatorname{tr} \frac{1}{x - \Phi} \right\rangle. \tag{3.10}$$

We denote the planar limit of $g_s W_1$ by $W_{0,1}$, given by

$$W_{0,1}(x) = g_s \sum_{i=1}^{N} \frac{1}{x - \lambda_i} \tag{3.11}$$

where $\{\lambda_i\}_{i=1,\ldots,N}$ satisfy (3.9). Since g_s is $\mathscr{O}(1/N)$ in the limit, and the sum is $\mathscr{O}(N)$, $W_{0,1}(x)$ is $\mathscr{O}(1)$. By deriving a quadratic equation for $W_{0,1}(x)$, one easily shows that in the large N limit $W_{0,1}(x)$ is given by

$$W_{0,1}(x) = \frac{1}{2}(V'(x) - \sqrt{(V'(x))^2 - 4f(x)}) \tag{3.12}$$

where the quantity $f(x)$ is a degree $d-2$ polynomial defined by

$$f(x) = g_s \sum_{i=1}^{N} \frac{V'(x) - V'(\lambda_i)}{x - \lambda_i}. \tag{3.13}$$

$W_{0,1}(x)$ has branch point singularities where the polynomial inside the square root vanishes. This polynomial is of degree $2d$, so in the generic situation one finds $2d$ branch points which may be joined by a choice of d cuts. This is the so-called multicut solution to the matrix model. To complete the solution of the model, the $d-1$ parameters in $f(x)$ must be determined precisely. We will return to this point momentarily.

It is useful to define the quantity

$$-y(x) = \frac{1}{2}V'(x) - W_{0,1}(x). \tag{3.14}$$

Up to a factor of 2 and a sign this is the effective force felt by a probe eigenvalue at the position x in the background of the equilibrium configuration. Then $y(x)$ satisfies the algebraic equation

$$y^2 - \frac{1}{4}(V'(x))^2 + f(x) = 0. \tag{3.15}$$

This defines an algebraic curve for $(x, y) \in \mathbb{C} \times \mathbb{C}$ known as the spectral curve because the analytic structure of the quantity $y(x)$ encodes the spectrum of eigenvalues in the large N limit. Strictly speaking, the spectral curve of the matrix model is a smooth Riemann surface determined by the normalization of the possibly singular curve defined above.

The Spectral Curve and its Moduli

The spectral curve is hyperelliptic of genus $g \leq d-1$. In the generic situation, $g = d-1$, but in general the curve may form singularities which decrease the genus by pinching cycles. The geometric picture coming from the standard procedure of making the Riemann surface for $y(x)$ by gluing sheets along cuts is shown in Figure 3.1 for the generic situation in a 3-cut matrix model (a curve of genus 2). Suppose there are d cuts and let A^i be the cycle wrapping the i-th cut clockwise. Suppose there are N_i eigenvalues at the i-th cut, so that the parameter $t^i = g_s N_i$ is finite as $N \to \infty$. These parameters are not independent, but satisfy the constraint $\sum_i t^i = t$, where t is regarded as a

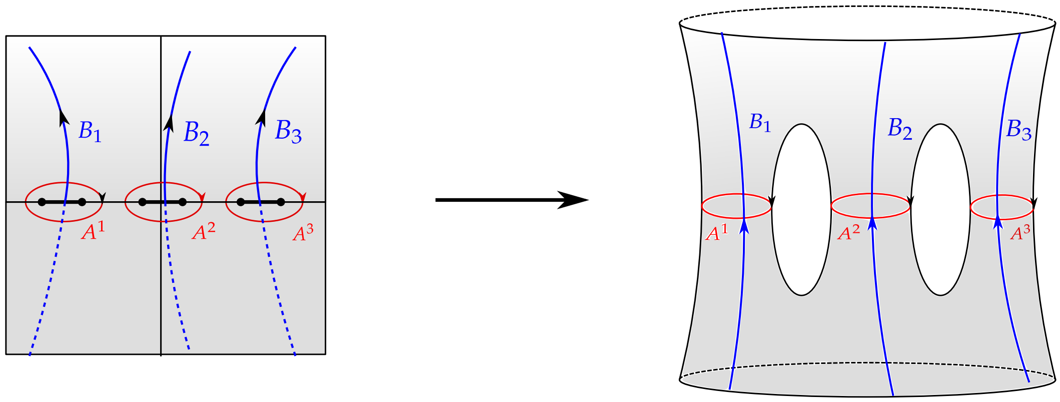

Figure 3.1: The A and B cycles in a 3-cut matrix model. The line segments denote the cuts of y, and the plane shown is the complex x-plane. The dotted lines on the B-cycles denote the passage to the other sheet, corresponding to the other branch of the square root appearing in $y(x)$. The surface on the right comes from gluing the two sheets together. Note the spectral curve is naturally noncompact because y has poles at infinity, but by adding two points at infinity it is compactified to a hyperelliptic curve of genus 2. The noncompact version is slightly more natural because the associated basis of relative homology cycles is better suited to the matrix model problem.

fixed external parameter. Then one has

$$t^i = \frac{1}{2\pi i}\oint_{A^i} y(x)dx \tag{3.16}$$

$$\frac{\partial F_0}{\partial t^i} = \int_{B_i} y(x)dx. \tag{3.17}$$

The cycle B_i is noncompact and is chosen to be canonically paired with the A-cycle, running from the cut to a point at infinity. F_0 is the planar free energy of the matrix model. We are not precise with the signs, 2π factors, etc in these equations. These type of equations are known as special geometry relations and are ubiquitous in string theory and supersymmetric gauge theory [8], [51]. The parameters t^i are known as the filling fractions and there are $d-1$ independent ones. Fixing them to prescribed values determines the form of $f(x)$ via the above relation.

The special geometry relations determine the solution of the matrix model–F_0 is the sum of all planar Feynman diagrams–purely in terms of the geometry of the spectral curve. These equations are closely analogous to the Seiberg-Witten geometry of $\mathcal{N}=2$ supersymmetric gauge theories, and it will be shown in a later chapter that this analogy is quite literal.

Reformulation: Loop Equations

With the hope of generalizing some of these constructions to more general quantum field theories, we reformulate this discussion in terms of Dyson-Schwinger equations for the matrix model known as loop equations. As usual with Dyson-Schwinger equations, the loop equations arise by integration by parts in the path integral. One begins with (implied sum on i, j)

$$0=\int[d\Phi]\frac{\partial}{\partial\Phi_{ij}}\Big(e^{-\frac{1}{g_s}\operatorname{tr}V(\Phi)}(\Phi^{\mu})_{ij}\Big). \tag{3.18}$$

One can then sum this relation over μ weighted by $x^{-\mu-1}$ to obtain the following equation for the generating function $W_1(x)$ introduced previously:

$$g_s^2\Big\langle\Big(\operatorname{tr}\frac{1}{x-\Phi}\Big)^2\Big\rangle=g_sV'(x)W_1(x)-f(x). \tag{3.19}$$

This is known as the loop equation. The object $f(x)$ is now defined by

$$f(x)=g_s\Big\langle\operatorname{tr}\frac{V'(x)-V'(\Phi)}{x-\Phi}\Big\rangle. \tag{3.20}$$

In the planar limit this reduces to the definition of the previous section. If one introduces the operator

$$Y(x)=g_s\operatorname{tr}\Big(\frac{1}{x-\Phi}\Big)-\frac{1}{2}V'(x) \tag{3.21}$$

then the loop equation may be interpreted as the statement that $\langle Y(x)^2 \rangle$ is a polynomial in x. In the planar limit, due to large N factorization one has $\langle Y(x)^2 \rangle \to \langle Y(x) \rangle^2$ and in this way one recovers the spectral curve of the previous section. It is important to note that the loop equations hold independent of any large N limit, and give rise to the spectral curve only as an effective, emergent object in this limit. The loop equations can also be extended to an infinite family of equations which allow one to recursively determine the $1/N^2$ corrections to the matrix model to any desired order of the large N expansion via a framework known as the topological recursion, but we will not elaborate on this further–see for example [14], [16].

The point of this analysis is that the Dyson-Schwinger equations of the matrix model can be recast using an operator $Y(x)$ which has poles when x is an eigenvalue of the random matrix Φ, and organized as the statement that $\langle Y(x)^2 \rangle$ has no poles in x.

Application to Random Partitions

The localization technique for supersymmetric gauge theories, to be explained in subsequent chapters, reduces partition functions of these theories to sums over discrete objects known as Young diagrams or partitions. In this section we illustrate that these sums over partitions can sometimes be rewritten as the partition function of a matrix model, which allows one to directly apply the machinery developed thus far to analyze the limiting behavior of such sums. This will be applied in later chapters to arrive at the Seiberg-Witten geometry of a large class of gauge theories.

A partition λ is defined as a nonincreasing sequence of positive integers $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$, and can be represented by a Young diagram. The quantity $|\lambda|$ is the number of boxes in the Young diagram, and is given by

$$|\lambda| = \sum_i \lambda_i. \tag{3.22}$$

The length of partition $\ell(\lambda)$ is the number of nonzero integers in the sequence; we consider partitions such that $\ell(\lambda) \leq N$, and take $N \to \infty$ at the end. It is customary to define the related sequence $h_i = \lambda_i - i + N$, which satisfies $h_1 > h_2 > ... > h_N \geq 0$. Geometrically, if one puts the Young diagram at 45° to horizontal, the h_i represent the x-coordinates of the outermost boxes (see Figure 3.2).

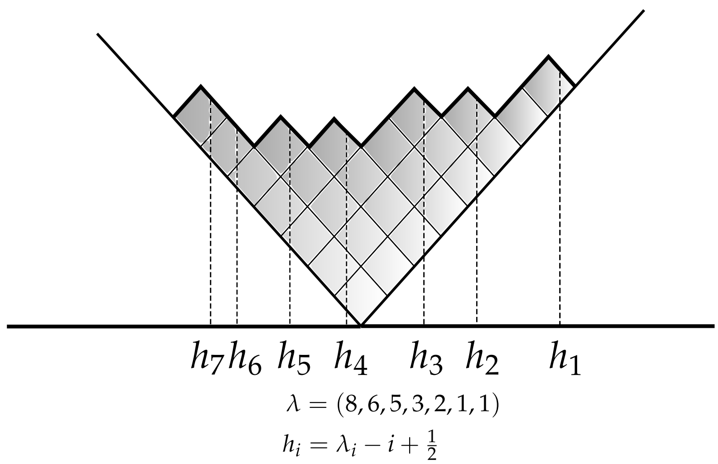

Figure 3.2: Young diagram corresponding to partition $\lambda = (8,6,5,3,2,1,1)$, drawn in the "Russian convention" along the diagonal. The h_i that appear in the figure differ from the h_i in the problem statement by some trivial translation by $N - \frac{1}{2}$. The piecewise linear curve in bold along the boundary of the diagram is known as the profile of the partition λ. In the limit of large partitions, after some rescaling it converges to a more general (no longer piecewise linear) Lipschitz function on an interval.

The partitions with k boxes index the irreducible representations of the symmetric group S_k [17]. The partition λ corresponding to the sequence $\{h_i\}_{i=1,\ldots,N}$ in turn corresponds to an irreducible representation of dimension

$$\dim \lambda = |\lambda|! \frac{\prod_{i<j}(h_i - h_j)}{\prod_i h_i!}. \tag{3.23}$$

It is a standard result, valid for any finite group G, that

$$|G| = \sum_R (\dim R)^2 \tag{3.24}$$

where $|G|$ denotes the number of elements of G and R runs over the set of irreducible representations of G. Therefore the quantity

$$\mathscr{P}(\lambda) = \frac{(\dim \lambda)^2}{k!} \tag{3.25}$$

defines a normalized measure on the set of partitions with k boxes, known as the Plancherel measure. Consider the partition function for the “Poissonized" Plancherel measure on partitions of length at most N:

$$Z(g_s, N) = \sum_{k=0}^{\infty} \frac{g_s^{-2k}}{k!} \sum_{|\lambda|=k, \ell(\lambda)\leq N} \mathscr{P}(\lambda). \tag{3.26}$$

It is clear that the $N \to \infty$ limit of Z with fixed g_s is simply e^{1/g_s^2}, since the Plancherel measure is normalized. In order to find something nontrivial, one must consider a limit with $N \to \infty$ and $g_s \to 0$ simultaneously, motivating the analogy with matrix models.

Consider the matrix model-like partition function (expressed as an eigenvalue integral)

$$Z(g_s, N) = \frac{1}{N!(2\pi i)^N} \oint_{\gamma^N} d^N x \prod_{i<j} (x_i - x_j)^2 \prod_{i=1}^{N} \varphi\left(\frac{x_i}{g_s}\right) \tag{3.27}$$

where

$$\varphi(x) = \frac{\Gamma(-x)}{\Gamma(x+1)} e^{i\pi x} g_s^{-2x-1}. \tag{3.28}$$

The integral above is taken for each eigenvalue x_i lying along a contour γ which encircles the positive real axis. The integral can therefore be evaluated by residues. Note that $\Gamma(x+1)$ is nowhere-vanishing on the positive real axis, so the only poles of $\varphi(x)$ come from the poles of

$\Gamma(-x)$. These occur at each nonnegative integer $n \geq 0$, and we recall the asymptotics

$$\Gamma(-n+\varepsilon) \sim \frac{(-1)^n}{n!}\frac{1}{\varepsilon} \tag{3.29}$$

for $\varepsilon \to 0$. Evaluating the integral by residues one finds

$$\begin{aligned} Z(g_s,N) &= \frac{g_s^{N(N-1)}}{N!}\sum_{h_i \geq 0}\prod_{i<j}(h_i-h_j)^2\prod_i \frac{g_s^{-2h_i}}{(h_i!)^2} \\ &= g_s^{N^2-N}\sum_{h_1>h_2>\ldots>h_N\geq 0}\frac{\prod_{i<j}(h_i-h_j)^2}{\prod_i (h_i!)^2} g_s^{-2\sum_i h_i}. \end{aligned} \tag{3.30}$$

To go from the first to second line, we have noted that we have invariance under the symmetric group S_N and used this to choose an ordering of the h's so that the sum coincides with a sum over partitions–note that terms with $h_i = h_j$ with $i \neq j$ do not contribute because of the Vandermonde determinant. Now we recall $h_i = \lambda_i - i + N$, so

$$\sum_i h_i = |\lambda| - \frac{N(N+1)}{2} + N^2 = |\lambda| + \frac{N(N-1)}{2}. \tag{3.31}$$

The integral is then

$$\begin{aligned} Z(g_s,N) &= \sum_{\ell(\lambda)\leq N} g_s^{-2|\lambda|}\frac{\prod_{i<j}(h_i-h_j)^2}{\prod_i (h_i!)^2} \\ &= \sum_{\ell(\lambda)\leq N}\frac{g_s^{-2|\lambda|}}{|\lambda|!}\mathscr{P}(\lambda). \end{aligned} \tag{3.32}$$

This reproduces the partition function defined via the Poissonized Plancherel measure. The matrix integral representation (3.27) provides a method for extracting the interesting asymptotics of the Plancherel measure.

To determine more precisely the scaling limit encoding these interesting asymptotics, note that for large N, Z is a sum of terms

$$\frac{g_s^{-2k}}{k!} = \exp(-2k\log g_s - \log k!). \tag{3.33}$$

To be accessible via the matrix model, this sum should be dominated by a saddle point as $g_s \to 0$. Clearly, for the terms to balance k must be large and we may use the Stirling formula $\log k! \sim k\log k - k$, and we see a saddle point exists provided

$$k \sim \frac{1}{g_s^2}. \tag{3.34}$$

Recalling that in terms of the partition λ, $k = |\lambda|$, which is the area of the region enclosed by the Young diagram (in units where the boxes have unit length), we see that the partition is becoming large with an area that grows like g_s^{-2} for $g_s \to 0$. This means that its length goes like $\ell(\lambda) \sim 1/g_s$, so the large N 't Hooft asymptotics of the matrix model captures the regime where the sum is dominated by large Young diagrams, and the saddle point dominating the sum is referred to as the limit shape. The 't Hooft parameter $t = g_s N$ just sets the units of length of the limiting partition. It may be set to any convenient value. We will choose such a value shortly. To find the precise form of the limit shape, we evaluate the matrix model for $Z(g_s, N)$ in the 't Hooft large N limit.

Note that the potential for the matrix model is given by $\sum_i \log\varphi\left(\frac{x_i}{g_s}\right)$, and we only care about the leading piece as $g_s \to 0$. Since

$$\varphi(x) = \frac{\Gamma(-x)}{\Gamma(x+1)} e^{i\pi x} g_s^{-2x-1} \tag{3.35}$$

the $g_s \to 0$ behavior can be obtained from the Stirling formula:

$$\Gamma(x+1) \sim x^x e^{-x}\sqrt{2\pi x} \tag{3.36}$$

for $x \to \infty$. Applying this formula, one finds

$$\varphi\left(\frac{x}{g_s}\right) \sim \exp\left\{-\frac{2x}{g_s}\log x + \frac{2x}{g_s} + \ldots\right\} \tag{3.37}$$

where $\ldots$ denotes terms less singular in g_s. This means that the $g_s \to 0$ asymptotics of Z is controlled by the matrix model

$$\int d^N x \prod_{i<j}(x_i - x_j)^2 e^{-\frac{1}{g_s}\sum_i V(x_i)} \tag{3.38}$$

where the potential is given by $V(x) = 2x\log x - 2x$, so that $V'(x) = 2\log x$. To solve the matrix model, we introduce the planar resolvent, expressed in terms of an effective eigenvalue density (a continuous function with compact support) $\rho(y)$ as

$$W_{0,1}(x) = g_s \sum_i \frac{1}{x - x_i} \to t\int dy \frac{\rho(y)}{x-y}. \tag{3.39}$$

Here, $\{x_i\}_{i=1,\ldots,N}$ denotes a set of x_i solving the saddle point equation. We have also indicated the large N limit of this expression, where the eigenvalues coalesce into a cut with some density $\rho(x)$ that will be presently determined. The resolvent clearly has asymptotics $W_{0,1}(x) \sim \frac{t}{x}$ for $x \to \infty$, and is analytic in the x-plane with a branch cut singularity along the support of the eigenvalue density.

The saddle point equation may be expressed in terms of the behavior of the resolvent across its cut as (see for example [29], [15])

$$W_{0,1}(x + i0^+) + W_{0,1}(x + i0^-) = 2\log x. \tag{3.40}$$

In this equation, x must be taken to lie along a cut of $W_{0,1}$. The potential $V(x)$ has a unique minimum at $x = 1$, so we expect a one-cut solution. To solve this equation, we note that if $V(x)$ were a polynomial, the above equation would state that $W_{0,1}(x)$ was single-valued on a two-sheeted ramified covering of the x-plane, so to this end we introduce a coordinate z on such a cover via the Joukowsky map:

$$x(z) = \alpha + \frac{\beta}{2}\left(z + \frac{1}{z}\right). \tag{3.41}$$

The sheets of the covering are exchanged by the map $z \mapsto 1/z$. The parameters α and β will be determined by the solution of the saddle point equation; they characterize the location of the cut occupied by the eigenvalues. In terms of z, $W_{0,1}(z) \sim \frac{2t}{\beta z}$ as $z \to \infty$. An obvious *ansatz* for $W_{0,1}(z)$ with the correct asymptotics and singular behavior is

$$W_{0,1}(z) = 2\log\left(1 + \frac{t}{\beta z}\right). \tag{3.42}$$

The saddle point equation (3.40) now determines α and β–substituting the *ansatz* yields:

$$\begin{aligned} \alpha &= 1 + \frac{t^2}{\beta^2} \\ \frac{\beta}{2} &= \frac{t}{\beta}. \end{aligned} \tag{3.43}$$

These equations have the solution $\beta = \sqrt{2t}$, $\alpha = 1 + \frac{t}{2}$, and from this we determine that the cut has endpoints $a = 1 + \frac{t}{2} - \sqrt{2t}$ and $b = 1 + \frac{t}{2} + \sqrt{2t}$. This completes the solution of the matrix model; for definiteness the resolvent is given as

$$W_{0,1}(z) = 2\log\left(1 + \sqrt{\frac{t}{2}}\frac{1}{z}\right). \tag{3.44}$$

To find the limit shape of the large partition, recall that the eigenvalues x_i are related to the variables h_i by $x_i = g_s h_i$. The profile of the partition as a function of the index i is given by $\lambda_i = h_i + i - N$, and we want the rescaled version $\tilde{\lambda}_i = x_i + g_s i - t$. To make sense of $g_s i$ we note the following: the index i as a function of x is simply the number of h_i, or equivalently the number of eigenvalues, which are to the right of that given value of x (the reader is encouraged to revisit Figure 3.2 if this is unclear).

Since, in terms of the individual eigenvalues, the density is given by $\rho(x) = \frac{1}{N}\sum_i \delta(x - x_i)$, the index $i(x)$ may be calculated as

$$i(x) = N\int_x^b \rho(x')dx' \tag{3.45}$$

where b is the upper endpoint of the cut. This continues to hold true in the $g_s \to 0$ limit, and one has for $\tilde{\lambda}_i$:

$$\tilde{\lambda}_{i(x)} = x - t + t\int_x^b \rho(x')dx'. \tag{3.46}$$

In principle, the relation $i(x)$ could be inverted and then the above equation could be used to plot λ as a function of $g_s i$. However, it is much more illuminating to consider the partition rotated by $\frac{\pi}{4}$, which is given by $\tilde{\lambda}_i + g_s i$ plotted against $\tilde{\lambda}_i - g_s i$–to this end, introduce the variable $u = \tilde{\lambda}_i - g_s i = x - t$ and consider plotting

$$\tilde{\lambda}_i + g_s i = x - t + 2t\int_x^b \rho(x')dx' \tag{3.47}$$

as a function of u. This will give the limit profile $f(u)$, that is $f(u) = \tilde{\lambda}_i(u) + g_s i(u)$.

To finish the calculation and determine $f(u)$ in closed form, all that remains is to determine the eigenvalue density $\rho(x)$. $\rho(x)$ is given by the discontinuity in $W_{0,1}(x)$ across its cut in the x-plane, so we must first determine $W_{0,1}$ as a function of x and then calculate its discontinuity.

The first step is simple, as one simply inverts the map $x(z)$ above to obtain $z(x)$:

$$\frac{1}{z} = \frac{1}{\beta}(x - \alpha - \sqrt{(x-\alpha)^2 - \beta^2}) \tag{3.48}$$

so we find

$$W_{0,1}(x) = 2\log\left(1 + \frac{1}{2}\left(x - 1 - \frac{t}{2} - \sqrt{\left(x - 1 - \frac{t}{2}\right)^2 - 2t}\right)\right). \tag{3.49}$$

Recall that t plays essentially no role in the problem other than setting an overall scale, so we may set it to a convenient value, say $t = 2$, so that the resolvent simplifies to

$$W_{0,1}(x) = 2\log\left(\frac{x}{2} - \frac{1}{2}\sqrt{x^2 - 4x}\right). \tag{3.50}$$

With this value of t, the cut is at $x \in [0,4] \subset \mathbb{R}$. We may compute the eigenvalue density from the discontinuity as

$$\rho(x) = \frac{1}{\pi}\operatorname{Im}\log\left(\frac{x}{2} + \frac{i}{2}\sqrt{4x - x^2}\right) = \frac{1}{\pi}\arctan\left(\frac{\sqrt{4x - x^2}}{x}\right) = \frac{1}{\pi}\arcsin\left(\frac{\sqrt{4-x}}{2}\right). \tag{3.51}$$

To further simplify this we can use some trigonometric identities

$$\begin{aligned}\rho(x) &= \frac{1}{\pi}\arcsin\left(\frac{\sqrt{4-x}}{2}\right)\\ &= \frac{1}{2\pi}\arccos\left(\cos\left(2\arcsin\left(\frac{\sqrt{4-x}}{2}\right)\right)\right)\\ &= \frac{1}{2\pi}\arccos\left(\frac{x-2}{2}\right)\\ &= \frac{1}{4} - \frac{1}{2\pi}\arcsin\left(\frac{x-2}{2}\right).\end{aligned} \tag{3.52}$$

From this we easily compute

$$g_s i(x) = 2\int_x^4 \rho(x')dx' = \frac{2-x}{2} + \frac{x-2}{\pi}\arcsin\left(\frac{x-2}{2}\right) + \frac{1}{\pi}\sqrt{4-(x-2)^2}. \tag{3.53}$$

From this we conclude, recalling $u = x - 2$ and $f(u) = u + 2g_s i(u)$:

$$f(u) = \frac{2}{\pi}\left(\sqrt{4-u^2} + u\arcsin\left(\frac{u}{2}\right)\right). \tag{3.54}$$

This is the famous Logan-Shepp-Vershik-Kerov profile [27], [59]. The point of relevance of this construction is that sums over random partitions with Plancherel or Plancherel-like measures tend to be dominated by limit shapes, which are large partitions that can be characterized by their limiting profile function. This profile function gives rise to a Riemann surface, in this construction it is simply the spectral curve of the underlying matrix model. In later chapters this will be understood as the mechanism producing Seiberg-Witten geometry of gauge theories.

CHAPTER 4: SUPERSYMMETRIC GAUGE THEORIES AND TOPOLOGICAL TWIST

We now define the gauge theories to be studied for the remainder of this thesis. They have $\mathcal{N} = 2$ supersymmetry in four dimensions, that is, eight real supercharges. Supersymmetric theories are considered solely for convenience because it is possible to perform exact computations in these theories. The present chapter will deal primarily with formal aspects of these theories and their Lagrangian descriptions, while the subsequent chapter will explain how to solve them using the localization technique. We begin with pure $\mathcal{N} = 2$ supersymmetric Yang-Mills theory, then explain the topological twisting procedure, and then explain how to produce a large class of $\mathcal{N} = 2$ theories by taking orbifold projections of the $\mathcal{N} = 4$ supersymmetric Yang-Mills theory. We neglect factors of $2pi, i$ etc in many of the formulas in this chapter because the explicit forms of the Lagrangians are not relevant for the subsequent analysis; the purpose of this chapter is merely to provide background for readers who are familiar with quantum field theory but less familiar with the intricacies of supersymmetry. We do assume a basic familiarity with supersymmetry, for an introduction to see for example [53], [7].

Pure $\mathcal{N} = 2$ super-Yang-Mills

The simplest way to construct supersymmetric gauge theories with extended supersymmetry is to start with a minimal supersymmetric theory in a higher dimension, and dimensionally reduce (i.e. compactify on a torus of zero size) to the dimension of interest.

The minimal supersymmetric Yang-Mills Lagrangian may be written as (the following discussion takes formulas from [7])

$$\mathscr{L} = -\frac{1}{4g_{YM}^2}\operatorname{tr} F_{\mu\nu}F^{\mu\nu} + \frac{i}{2g_{YM}^2}\operatorname{tr}\lambda^a\Gamma^\mu_{ab}D_\mu\lambda^b. \tag{4.1}$$

Here, λ^a denotes a fermionic field in the minimal real spinor representation S of the group $\mathrm{Spin}(1,d-1)$ of spacetime symmetries in dimension d and in the adjoint representation of the gauge group, which we take to be $SU(n)$ for simplicity. Γ^μ_{ab} denote the coefficients of the essentially unique symmetric $\mathrm{Spin}(1,d-1)$-equivariant morphism of representations $\Gamma : S\otimes S \to V$, where V denotes the defining representation of $SO(1,d-1)$. $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$ is the field strength of the gauge field A_μ. tr is taken over the gauge indices, and D_μ denotes the covariant derivative in the adjoint representation.

This Lagrangian is supersymmetric in $d = 3,4,6,10$ with supersymmetry variations parameterized by a spinor η^a:

$$\begin{aligned} \delta A_\mu &= \eta^a\Gamma^\mu_{ab}\lambda^b \\ \delta\lambda^a &= \frac{1}{2}\eta^b\widetilde{\Gamma}^{\mu ac}\Gamma^\nu_{cb}F_{\mu\nu}. \end{aligned} \tag{4.2}$$

The objects $\widetilde{\Gamma}^{\mu ac}$ are the coefficients of the dual morphism $\widetilde{\Gamma} : S^*\otimes S^* \to V$, which satisfy the Clifford relation $\Gamma^\mu_{ac}\widetilde{\Gamma}^{\nu cb} + \Gamma^\nu_{ac}\widetilde{\Gamma}^{\mu cb} = 2g^{\mu\nu}\delta^b_a$, where $g^{\mu\nu}$ is the inverse of the Minkowski metric. Associated to this symmetry is a conserved charge Q_a which transforms as a spinor.

In six dimensions, the minimal real spinor has eight real degrees of freedom, and therefore the dimensional reduction of the above theory will produce a theory in four dimensions with eight supercharges, that is, $\mathscr{N} = 2$ supersymmetry, which is simply $\mathscr{N} = 2$ supersymmetric Yang-Mills theory.

There is a somewhat subtle point when dealing with real spinors in $d = 5+1$, owing to the fact that the irreducible complex spinor representations of $\mathrm{Spin}(1,5)$ are pseudoreal. The reality condition on fermions can only be imposed after doubling the fields, that is to say the fermions λ in $d = 5+1$ carry indices as λ^{aI}, where a is the index for the (say) left-handed chiral spinor representation, ranging from $a = 1,\ldots,4$ and $I = 1,2$. The reality condition then reads ($j^a{}_b$ are the coefficients of the quaternionic structure of the spinor representation)

$$(\lambda^{aI})^* = j^a{}_b \varepsilon_{IJ} \lambda^{bJ}. \tag{4.3}$$

Note that this equation is $\mathrm{Spin}(1,5) \times SU(2)$-equivariant, so that I can be thought of as an $SU(2)$ index. This is the so-called symplectic Majorana condition. It is necessary to consider the symplectic Majorana spinors in what follows since the general supersymmetric action discussed above is written for a minimal real spinor.

Upon reduction to $d = 4$, the six-dimensional supercharge descends to a pair Q^I_α of four-dimensional supercharges, such that $(Q^I_\alpha)^\dagger = Q^\dagger_{I\dot{\alpha}}$ (using the standard dotted/undotted notation for four-dimensional chiral spinors). The six-dimensional supersymmetry algebra reduces to the $\mathscr{N} = 2$ supersymmetry algebra in four dimensions. The momentum $P_4 + iP_5$ descends to the central charge Z of four-dimensional supersymmetry, and the BPS states in four dimensions correspond to the massless states in six dimensions.

With the details of the fermionic fields addressed, the field content of the pure $\mathscr{N} = 2$ supersymmetric gauge theory is simple to analyze. The six dimensional gauge field A_M reduces to a four-dimensional gauge field A_μ and two scalars A_4,A_5, which may be combined into a single complex scalar ϕ in the adjoint representation. The six-dimensional spinor λ^{aI} reduces to a pair of four-dimensional Weyl fermions $\lambda^{\alpha I}$, also in the adjoint representation. Their adjoints $\bar{\lambda}^{\dot{\alpha}}_I$ are right-handed Weyl fermions. The group $SU(2)$ acting on the I index is a global symmetry of $\mathscr{N} = 2$ super-Yang-Mills theory known as R-symmetry, which we denote by $SU(2)_I$. The field content is a representation of $\mathscr{N}=2$ supersymmetry known as the vector multiplet.

For completeness, the Lagrangian of the $\mathscr{N}=2$ supersymmetric Yang-Mills theory is (ignoring irrelevant factors including the coupling constant g_{YM}^2; interested readers may perform the dimensional reduction explicitly to find the factors):

$$\begin{aligned}\mathscr{L}\sim -\frac{1}{4}\operatorname{tr}F_{\mu\nu}F^{\mu\nu}+\operatorname{tr}D_\mu\phi^\dagger D^\mu\phi-\operatorname{tr}\left[\phi,\phi^\dagger\right]^2\\ -i\operatorname{tr}\bar{\lambda}_I^{\dot{\beta}}\sigma^\mu_{\alpha\dot{\beta}}D_\mu\lambda^{\alpha I}+i\varepsilon_{\alpha\beta}\varepsilon_{IJ}\operatorname{tr}\lambda^{\alpha I}\left[\phi^\dagger,\lambda^{\beta J}\right]+c.c.\end{aligned} \tag{4.4}$$

The objects $\sigma^\mu_{\alpha\dot{\beta}}$ are well-known, they are simply the gamma matrices acting on chiral spinors. In addition to this piece, the well-known θ term should be added to the action, most conveniently expressed as

$$\frac{\theta}{8\pi^2}\int\operatorname{tr}(F_A\wedge F_A) \tag{4.5}$$

where $F_A=dA+A\wedge A$ is the 2-form representing the field strength of the gauge field A. In supersymmetric theories, it is useful to introduce the complexified coupling

$$\tau=\frac{\theta}{2\pi}+\frac{4\pi i}{g_{YM}^2}. \tag{4.6}$$

Many quantities of interest will depend holomorphically on τ.

Topological Twist

The key to performing exact calculations in $\mathscr{N}=2$ supersymmetric theories is the topological twisting procedure introduced by Witten [60]. The procedure of topological twisting can be motivated by the desire to define a supersymmetric theory on a curved spacetime background–without modification, the theory will be supersymmetric on a curved manifold only if the manifold admits covariantly constant spinors. The procedure of twisting alters the way that the theory couples to gravity so that there is a scalar supercharge on an arbitrary curved background.

Concretely, the way this is done is as follows. Coupling to gravity can be thought of as gauging the Lorentz group. The Lorentz group for a Euclidean theory is $\mathrm{Spin}(4) = SU(2)_L \times SU(2)_R$. For $\mathcal{N} = 2$ super-Yang-Mills, we also have the R-symmetry $SU(2)_I$ and may consider the action of the full group $SU(2)_L \times SU(2)_R \times SU(2)_I$.

The twisted theory is produced by defining $SU(2)'_R$ to be the diagonal subgroup of $SU(2)_R \times SU(2)_I$ and treating the group $SU(2)_L \times SU(2)'_R$ as the “Lorentz group" of the theory. Practically, this means replacing the index I of R-symmetry by $\dot{\alpha}$ of $SU(2)_R$. In flat space, where the Lorentz group is not gauged, this is just a change in notation that reorganizes the same field content. However, the new organization is quite powerful for exact calculations.

It is useful to understand how the various fields in the theory behave after twisting. The left-handed fermions $\lambda^{\alpha I}$ are replaced with the object $\lambda^{\alpha\dot{\beta}}$, which is an object in the $(\mathbf{\frac{1}{2}}, \mathbf{\frac{1}{2}})$ representation of $SU(2) \times SU(2)$, in other words a four-vector ψ_μ.

Likewise, the right-handed fermions $\bar{\lambda}_{\dot{\alpha}I}$ become an object in the $(\mathbf{0}, \mathbf{\frac{1}{2}}) \otimes (\mathbf{0}, \mathbf{\frac{1}{2}}) = (\mathbf{0}, \mathbf{0}) \oplus (\mathbf{0}, \mathbf{1})$ representation of $SU(2) \times SU(2)$, which is the same as a scalar η together with a self-dual[1] two-form $\chi^+_{\mu\nu}$. Note that the twisting procedure takes place in Euclidean signature, where all fields are naturally complex and the reality conditions of Minkowski signature must be relaxed–thus, $\bar{\lambda}_{\dot{\alpha}I}$ is no longer the complex conjugate of a left-handed fermionic field, it is regarded as an independent entity.

---

[1]Depending on one’s convention for the Lorentz group, self-dual may be replaced with anti-self-dual here. The two choices are related by parity, which exchanges $SU(2)_L$ and $SU(2)_R$.

Similarly, the supercharges of the theory are reorganized as

$$Q^I_\alpha, Q^\dagger_{\dot\alpha I} \to Q, G_\mu, Q^+_{\mu\nu}. \tag{4.7}$$

The scalar supercharge generates a fermionic symmetry δ of the field space which satisfies $\delta^2 = 0$ up to gauge transformation. The action of the theory may be written as

$$S = (\text{topological term}) + \delta(...). \tag{4.8}$$

Treating δ as an exterior derivative on field space, by usual integration by parts arguments the δ-exact terms decouple. This fact can be used to show that the correlation functions of the twisted theory on an arbitrary four-manifold produce the so-called Donaldson invariants [60], [9].

The supersymmetry of the twisted theory is most conveniently represented by introducing an auxiliary self-dual two-form field H^+, in which case the supersymmetry variations under the scalar supercharge δ take the form (all fields are written in differential forms notation in these equations, and we replace the notation $\phi^\dagger$ with $\bar\phi$):

$$\begin{array}{lll} \delta A = \psi & \delta\phi = 0 & \delta\chi^+ = H^+ \\ \delta\psi = D_A\phi & \delta\bar\phi = \eta & \delta H^+ = [\chi^+, \phi] \\ & \delta\eta = [\bar\phi, \phi] & \end{array} \tag{4.9}$$

The action of the theory may be rewritten in the twisted formalism as (we are again imprecise with numerical factors)

$$S = -\frac{2\pi i \tau}{8\pi^2}\int_{\mathbb{R}^4} \mathrm{tr}\, F_A \wedge F_A \\ + \delta\left(\int_{\mathbb{R}^4} -\mathrm{tr}\,\psi \wedge \star D_A \bar{\phi} + \mathrm{tr}\,\eta \star [\bar{\phi},\phi] - \mathrm{tr}\,\chi^+ \wedge \left(F_A^+ - \frac{g_{YM}^2}{8} H^+\right)\right). \tag{4.10}$$

Note that all metric dependence in the action is δ-exact and hence decouples, which is responsible for the name topological twist. In this equation, F_A^+ denotes the self-dual part of the two-form F_A. After integrating out the auxiliary field H^+, this action agrees with that above for $\mathscr{N} = 2$ super-Yang-Mills when both theories are defined in flat space, up to the changes in notation due to the twisting.

Consulting the discussion in, e.g. [5] one recognizes the supersymmetry variation δ as representing the equivariant de Rham differential on the space $\mathscr{A}$ of all connections on an $SU(n)$ principal bundle over $\mathbb{R}^4$, equivariant with respect to the infinite-dimensional group $\mathscr{G}$ of gauge transformations (automorphisms of the principal bundle). The fields are organized into multiplets of the δ-symmetry as (A,ψ), $(\phi,\bar{\phi},\eta)$, and (χ^+,H^+), and one recognizes the δ-exact piece of the action as essentially representing a supersymmetric delta function supported on the locus $F_A^+ = 0$, modulo the action of the group $\mathscr{G}$. This is one way to define the moduli space of instantons in gauge theory, which is actually a finite dimensional space (strictly speaking this is true when gauge fields are considered within a fixed connected component of $\mathscr{A}$, that is, gauge fields have fixed second Chern class or instanton charge), so in this way the path integral reduces to a finite-dimensional integral.

Introduce the space

$$\mathscr{M}(k,n) = \{A \in \mathscr{A} | F_A^+ = 0,\ c_2 = k\}/\mathscr{G}_\infty \tag{4.11}$$

where $c_2 = \frac{1}{8\pi^2}\int \mathrm{tr}(F_A \wedge F_A)$ denotes the second Chern class of the gauge field configuration. The quotient is with respect to the group $\mathscr{G}_\infty$ of gauge transformations which approach 1 at infinity. This space is known as the moduli space of framed instantons. For an elementary introduction to instantons see [53].

We address one subtle point closely related to the framing of the instantons. Strictly speaking, $D_A\bar{\phi}$ is a generator of a transformation in $\mathscr{G}_\infty$ only if $\bar{\phi} \to 0$ at infinity. However, this is not necessarily the case. As is evident from $\delta(\mathrm{tr}\,\eta[\bar{\phi},\phi]) = \mathrm{tr}[\phi,\bar{\phi}]^2 + ...$, for the physical choice of integration contour over the scalar fields in the path integral with $\bar{\phi} = \phi^\dagger$ the vacua of the theory are at $\left[\phi,\phi^\dagger\right] = 0$, which are the set of ϕ diagonalizable by a gauge transformation (recall the gauge group is $SU(n)$ gauge theory for some n). Without loss of generality the asymptotic value of ϕ is set to $\phi_\infty = \mathrm{diag}(a_1,...,a_n)$ where a_α are complex numbers, $\alpha = 1,...,n$ which sum to zero. It is only the nonzero modes of ϕ which are integrated over in the path integral–the a_α stay as fixed parameters describing the vacuum of the theory (more mathematically, they characterize a boundary condition on the fields integrated over in the path integral). The notation $\mathbf{a} = (a_1,...,a_n)$ is convenient. These parameters are also referred to as Coulomb moduli. This means that the path integral represents a supersymmetric delta function along $F_A^+ = 0$, and implements a quotient only by the group $\mathscr{G}_\infty$. The zero modes of the scalar field become equivariant parameters (in the sense of equivariant cohomology [56], [23]) for the group G of constant gauge transformations acting nontrivially on $\mathscr{M}(k,n)$.

The path integral therefore reduces to an integral over the space $\mathscr{M}(k,n)$, where the integrand is determined by the surviving zero modes of the fields (introduce the notation $\mathfrak{q} = e^{2\pi i\tau}$ and recall that one must sum over all instanton numbers–note $F_A^+ = 0$ has no solutions for k negative):

$$Z(\mathbf{a};\mathfrak{q}) = \sum_{k\geq 0} \mathfrak{q}^k \int_{\mathscr{M}(k,n)} e^{-\delta(g(V(\bar{\phi}_\infty),\psi))}. \tag{4.12}$$

In this equation, g is the metric on $\mathscr{M}(k,n)$ inherited from the field theory action (which is the same as the one inherited from the natural metric on $\mathscr{A}$), $V(\bar{\phi}_\infty)$ is the vector field on $\mathscr{M}(k,n)$ descending from $D_A\bar{\phi}_\infty = [A, \bar{\phi}_\infty]$ that generates constant gauge transformations, and finally ψ is the zero mode of the field of the same name that has survived the projections. This reduces the problem of computing the gauge theory partition function to a problem of integrating over the moduli space of instantons. In the next chapter, the evaluation of these integrals will be explained.

ADE Quiver Gauge Theories

In this section, we briefly sketch a construction of a large class of $\mathscr{N}=2$ supersymmetric gauge theories which have their origin in $\mathscr{N}=4$ supersymmetric Yang-Mills theory, known as *ADE* quiver gauge theories. In fact, the natural home of these constructions is within string theory.

The starting point for the construction is the $\mathscr{N}=4$ supersymmetric Yang-Mills theory, which is obtained from dimensional reduction of the minimal supersymmetric theory in $d = 9+1$ to four dimensions. To analyze the field content, one restricts to the subgroup $\mathrm{Spin}(1,3)\times\mathrm{Spin}(6)\subset \mathrm{Spin}(1,9)$. Since $\mathrm{Spin}(6)\cong SU(4)$, the theory has an $SU(4)$ global symmetry, which is identified as the R-symmetry of $\mathscr{N}=4$ supersymmetry. The supercharges carry indices as Q^i_α, where $i = 1,\ldots,4$ is an index for $SU(4)$.

The vector multiplet for $\mathscr{N}=4$ supersymmetry is obtained by reduction of the vector multiplet from $9+1$ dimensions. The ten-dimensional gauge field becomes a four dimensional gauge field A and six scalars which transform as the vector of $\mathrm{Spin}(6)$, so the scalars organize into a two-index antisymmetric tensor of $SU(4)$ denoted ϕ^{ij}. The fermionic content of the theory consists of four Weyl fermions $\lambda^{\alpha i}$ transforming as the fundamental of $SU(4)$. All fields are in the adjoint representation of the gauge group, which we take to be $U(n)$ for simplicity.

To reduce the supersymmetry, one chooses a discrete subgroup $\Gamma\subset SU(4)$ and restricts to only the

Γ-invariant fields. It is useful to rescale the rank of the gauge group to consider the $U(n|\Gamma|)$ gauge theory. The centralizer of Γ inside of $SU(4)$ dictates how much supersymmetry is preserved. Fix a subgroup $\Gamma \subset SU(2) \subset SU(2) \times SU(2) \subset SU(4)$. This choice will preserve $SU(2)$ R-symmetry and therefore produce an $\mathcal{N} = 2$ supersymmetric theory.

Discrete subgroups $\Gamma \subset SU(2)$ have an *ADE* classification (for more detail on some of what follows, see for example [46]), so the resulting quiver gauge theories are referred to as the *ADE* quiver gauge theories. To obtain any interesting structure, the discrete symmetry Γ must be related to the gauge symmetry. A canonical way to do this is to realize the gauge group $U(n|\Gamma|)$ as acting on the vector space $\mathbb{C}^n \otimes \mathbb{C}^{|\Gamma|}$, and interpret the second factor as the regular representation of Γ. The regular representation decomposes into the irreducible representations of Γ as

$$\mathbb{C}^{|\Gamma|} = \bigoplus_{i \in \Gamma^\vee} \mathbb{C}^{a_i} \otimes R_i \tag{4.13}$$

where $a_i = \dim R_i$ is the dimension of the irreducible representation R_i, and $\Gamma^\vee$ denotes the set of irreducible representations of Γ. It is also important to note that, if T denotes the defining representation of $SU(2)$ regarded as a representation of Γ, then

$$T \otimes R_i = \bigoplus_{j \in \Gamma^\vee} \mathbb{C}^{A_{ij}} \otimes R_j \tag{4.14}$$

where A_{ij} is the adjacency matrix of the corresponding *ADE* affine Dynkin graph, where the *ADE* type is determined by Γ. The main case we consider in future sections is $\Gamma = \mathbb{Z}_{r+1}$, which leads to the affine Dynkin graph of type $\widehat{A}_r$. In what follows, we use in an essential way the McKay correspondence, which (among other things) identifies the set $\Gamma^\vee$ with the set of vertices of the affine Dynkin graph of the same *ADE* type as Γ.

With these notations, the fields decompose as follows. The gauge field A_μ transforms trivially under R-symmetry, so the projection onto the Γ-invariant part (i.e. the restriction to $\mathrm{End}_\Gamma(\mathbb{C}^n \otimes \mathbb{C}^{|\Gamma|})$) produces a set of gauge fields $A^{\mathbf{i}}_\mu$, where $\mathbf{i} \in \Gamma^\vee$ runs over the set of irreducible representations of Γ, or the vertices of the corresponding affine Dynkin graph. The gauge group is reduced to $\prod_{i\in\Gamma^\vee} U(na_i)$, so there is one unitary gauge group factor for each vertex of the corresponding Dynkin graph.

The scalars transform nontrivially under R-symmetry. Introduce indices (A,I), both taking the values $1,2$, associated to the subgroup $SU(2)\times SU(2) \subset SU(4)$. The index A will be "eaten" by the orbifold projection, while the index I survives as the $SU(2)$ R-symmetry index in the resulting theory. Then the scalars ϕ^{ij} decompose under $SU(2)\times SU(2)$ as two real scalars transforming trivially, and a complex field Q_{AI} which inherits the reality condition $(Q_{AI})^\dagger = \varepsilon^{AB}\varepsilon^{IJ}Q_{BJ}$ from the fact that ϕ^{ij} is in a real representation of $SU(4)$. The two real scalars may be combined into a complex scalar Φ, which decomposes exactly as the gauge field to a set $\Phi^{\mathbf{i}}$ of scalars labeled by the vertices of the corresponding ADE Dynkin graph. These are simply the scalars of the $\mathcal{N}=2$ vector multiplets built on the gauge fields $A^{\mathbf{i}}_\mu$.

The other scalars decompose according to (in fact there is an $SU(2)$ R-symmetry doublet worth of each field, but we suppress the index I because it is acted upon trivially by Γ, alternatively the $I=2$ fields are essentially the complex conjugates of the $I=1$ fields due to the reality condition):

$$\begin{aligned}\mathrm{Hom}_\Gamma(\mathbb{C}^n\otimes\mathbb{C}^{|\Gamma|}, T\otimes\mathbb{C}^n\otimes\mathbb{C}^{|\Gamma|}) &= \bigoplus_{i,j,k}\mathrm{Hom}_\Gamma(\mathbb{C}^{na_i}\otimes R_i, \mathbb{C}^{na_j A_{jk}}\otimes R_k)\\ &= \bigoplus_{i,j}\mathrm{Hom}(\mathbb{C}^{na_i},\mathbb{C}^{na_j})\otimes\mathbb{C}^{A_{ij}}.\end{aligned} \tag{4.15}$$

We have used the fact that A_{ij} is a symmetric matrix. Under the McKay correspondence, the nonzero entries of A correspond to the edges of the affine ADE Dynkin graph, such that A_{ij} connects vertices $\mathbf{i}$ and $\mathbf{j}$. Call the graph γ, its set of edges Edge_γ, and its set of vertices Vert_γ. Orient the edges such that each edge e has a source $s(e) = \mathbf{i} \in \mathrm{Vert}_\gamma$ and target $t(e) = \mathbf{j} \in \mathrm{Vert}_\gamma$. Then one has, for the decompositon of Q_{A1}:

$$\bigoplus_{e \in \mathrm{Edge}_\gamma} \mathrm{Hom}(\mathbb{C}^{na_{s(e)}}, \mathbb{C}^{na_{t(e)}}) \oplus \mathrm{Hom}(\mathbb{C}^{na_{t(e)}}, \mathbb{C}^{na_{s(e)}}). \tag{4.16}$$

Thus, to each edge of the Dynkin graph one has a pair of complex scalars $(Q_e, \widetilde{Q}_e)$, transforming in dual bi-fundamental representations of the gauge group–if Q_e connects vertices $\mathbf{i}$ and $\mathbf{j}$, and the graph is oriented such that $t(e) = \mathbf{j}$, then Q_e is in the fundamental of $U(na_j)$ and the anti-fundamental of $U(na_i)$, and vice versa for $\widetilde{Q}_e$. The reality condition on the original set of Q's implies that the pair

$$\begin{pmatrix} Q_e \\ -\widetilde{Q}_e^\dagger \end{pmatrix} \tag{4.17}$$

transforms in the doublet of $SU(2)$ R-symmetry. This is the bosonic field content of the so-called $\mathcal{N} = 2$ hypermultiplet, which plays the role of a matter field in the $\mathcal{N} = 2$ supersymmetric context.

The fermions are simple to analyze now that these cases have been worked out. The fundamental of $SU(4)$ decomposes into the direct sum of the fundamental of each $SU(2)$ factor of $SU(2) \times SU(2) \subset SU(4)$, so the fermions transform under $SU(2) \times SU(2)$ as $\lambda^{\alpha A}$, and $\lambda^{\alpha I}$. The fields $\lambda^{\alpha I}$, after orbifold projection become a set $\lambda_{\mathbf{i}}^{\alpha I}$, $\mathbf{i} \in \mathrm{Vert}_\gamma$ of vector multiplet fermions, transforming as the doublet of $SU(2)$ R-symmetry. The fields $\lambda^{\alpha A}$ decompose as the scalars to pairs of fields $(\psi_e^\alpha, \widetilde{\psi}_e^\alpha)$ for each edge of the Dynkin graph, transforming in dual bi-fundamental representations of the gauge group. These fields are the hypermultiplet fermions, and they are acted on trivially by R-symmetry.

The Lagrangian for $\mathcal{N} = 4$ supersymmetric Yang-Mills theory is, again neglecting numerical factors,

$$\mathscr{L} \sim -\frac{1}{4}\operatorname{tr} F_{\mu\nu}F^{\mu\nu} + \varepsilon_{ijkl}\operatorname{tr} D_\mu \phi^{ij} D^\mu \phi^{kl} - \varepsilon_{ijkl}\varepsilon_{mnpq}\operatorname{tr}\left[\phi^{ij},\phi^{mn}\right]\left[\phi^{kl},\phi^{pq}\right] \\ - i\operatorname{tr}\bar{\lambda}_i^{\dot\beta}\sigma^\mu_{\alpha\dot\beta}D_\mu\lambda^{\alpha i} + \varepsilon_{\alpha\beta}\varepsilon_{ijkl}\operatorname{tr}\lambda^{\alpha i}\left[\phi^{jk},\lambda^{\beta l}\right] + c.c. \tag{4.18}$$

The action for the ADE quiver gauge theories can be obtained by inserting the Γ-invariant fields into this expression, but it is not particularly illuminating. Suffice it to say that there are now complexified gauge couplings $\tau_{\mathbf{i}}$ as well as Coulomb moduli for each node of the graph, since each node has its own vector multiplet. The couplings between the vector multiplets are via the hypermultiplets in the bi-fundamental representations of the gauge group.

Since the field content of the ADE quiver gauge theories can be written purely in terms of the underlying Dynkin graph, ignoring the fact that it is an affine Dynkin graph of type ADE, it is possible to define quiver gauge theories for arbitrary oriented graphs. We will not pursue this, in fact the quiver gauge theories in Cartan type A are already quite nontrivial.

Quiver gauge theories associated to finite, rather than affine, Dynkin graphs can be obtained by setting the gauge couplings associated to certain nodes to zero, for example the quiver gauge theory of type A_1 can be obtained from the $\widehat{A}_2$ theory from decoupling the nodes associated to the 0 and 2 representations of the McKay dual group $\mathbb{Z}_3$, leaving behind a theory of a $\mathscr{N}=2$ vector multiplet with gauge group $U(n)$, coupled to n hypermultiplets in the fundamental and n hypermultiplets in the antifundamental representation. The flavor symmetry of the theory associated to the multiplicity of the hypermultiplets can be interpreted as a gauge symmetry with a nondynamical vector multiplet, and the associated mass terms of the hypermultiplets can be interpreted as the Coulomb moduli of the decoupled nodes. These features will resurface in the following chapter when we perform explicit calculations in these theories.

CHAPTER 5: INSTANTON MODULI SPACES AND LOCALIZATION

In the previous chapter it was demonstrated that, upon topological twisting, the path integral for pure $\mathcal{N} = 2$ supersymmetric Yang-Mills theory reduces to a finite-dimensional integral over the moduli space of instantons $\mathscr{M}(k,n)$. In the present chapter we explain how to use this fact to solve the theories, by introducing a certain compactification $\overline{\mathscr{M}}(k,n)$ of the instanton moduli space and deformation of the $\mathcal{N} = 2$ supersymmetric field theory which allows the resulting integrals to be evaluated by the Atiyah-Bott localization formula.

ADHM Construction of Instantons

Recall that Yang-Mills instantons (for an elementary introduction to instantons see [53]) are solutions to the self-duality equation

$$F_A^+ = 0 \tag{5.1}$$

for a gauge field A in $\mathbb{R}^4$ with second Chern class k, modulo the action of gauge symmetries which approach 1 at infinity. Atiyah, Drinfeld, Hitchin, and Manin [2] provided a method to characterize the space of solutions to this equation in terms of a set of matrices.

Dirac Zero Modes

To motivate the construction, recall the following simple fact: an instanton in a d-dimensional theory may be thought of as a soliton in a $(d+1)$-dimensional theory. To understand solitons, it is natural to probe their interaction with other particles in the theory. In particular, one can search for bound states of these solitons and Dirac fermions charged under the fundamental representation of

$U(n)$.

Concretely, what this means is one wishes to find L^2 normalizable solutions to the Dirac equation in four dimensional Euclidean spacetime in the background of an instanton, that is, spinors ψ valued in the vector bundle E associated with the fundamental representation, satisfying

$$\not{D}_A \psi = 0. \tag{5.2}$$

The bundle S of spinors splits as $S = S_+ \oplus S_-$, where $S_\pm$ are the bundles of positive and negative chirality spinors. After choosing a complex structure to identify $\mathbb{R}^4 \cong \mathbb{C}^2$, one may identify

$$\begin{aligned} S_+ &\cong (\Omega^{0,0} \oplus \Omega^{0,2}) \otimes \sqrt{K} \\ S_- &\cong \Omega^{0,1} \otimes \sqrt{K}. \end{aligned} \tag{5.3}$$

$\sqrt{K}$ denotes the square root of the canonical bundle of $\mathbb{C}^2$, and $\Omega^{p,q}$ denotes the bundle of forms with p holomorphic indices and q anti-holomorphic indices; thus $K = \Omega^{2,0}$. Since $\sqrt{K}$ is trivial, it can be omitted and spinors may be regarded as differential forms. With these identifications, one has $\not{D}_A = \sqrt{2}(\bar{\partial}_A - \bar{\partial}_A^\dagger)$.

Using this description, one readily shows that there are no nontrivial solutions to the Dirac equation in $(\Omega^{0,0} \oplus \Omega^{0,2}) \otimes E$. From an index theorem, one concludes that the dimension of the space of $\chi \in \Gamma(\Omega^{0,1} \otimes E)$ solving the Dirac equation in the background of instanton charge k is k. Let $K \cong \mathbb{C}^k$ denote this space, that is, K the kernel of $\not{D}_A$ acting on negative chirality spinors. Let $N \cong \mathbb{C}^n$ denote the fiber of the bundle E over the $\mathbb{S}^3$ at infinity; since A approaches pure gauge at infinity, E has a natural trivialization over $\mathbb{S}^3$ and thus all fibers may be identified with N. The claim of the ADHM construction is that K and N, together with some linear maps between them, allow one to completely reconstruct the instanton solution.

Write $(z_1, z_2) \in \mathbb{C}^2$ as the coordinates in $\mathbb{C}^2 \cong \mathbb{R}^4$; then there are natural maps $\Gamma(S \otimes E) \to \Gamma(S \otimes E)$ given by $\psi \mapsto z_\alpha \psi$ for $\alpha = 1,2$. These operators commute, however their projections to the subspace of spinors solving the Dirac equation may not. If $\Pi : \Gamma(S \otimes E) \to K$ denotes orthogonal projection, then we define the operators

$$B_\alpha := \Pi \circ z_\alpha \bigg|_K \in \mathrm{End}(K) \tag{5.4}$$

for $\alpha = 1,2$.

Let χ denote the row vector of size k whose entries are some orthonormal basis of solutions to the Dirac equation–note that these solutions are sections of $\Omega^{0,1} \otimes E$, so χ is an $n \times k$ matrix with entries valued in $\Omega^{0,1}$. Then the operators $B_1, B_2 \in \mathrm{End}(K) \cong \mathrm{Mat}_k(\mathbb{C})$ are (we rescale by a factor $\pi\sqrt{2}$ for later convenience)

$$B_\alpha = \frac{1}{\pi\sqrt{2}} \int_{\mathbb{R}^4} z_\alpha \chi^\dagger \wedge \star \chi \tag{5.5}$$

for $\alpha = 1,2$. Note that orthonormality means that

$$\int_{\mathbb{R}^4} \chi^\dagger \wedge \star \chi = \mathbb{1}_K. \tag{5.6}$$

Thus there is a natural action of $U(k)$ by $\chi \mapsto \chi U$ for $U \in U(k)$ which acts on the B's by conjugation. A basis-free way to write this is $U(K)$ the group of unitary transformations preserving the canonical hermitian structure on the vector space K.

To learn about N, consider the limit $r = \sqrt{|z_1|^2 + |z_2|^2} \to \infty$. The Dirac equation states, for the $(0,1)$ spinors, that $\bar{\partial}_A \chi = \bar{\partial}_A^\dagger \chi = 0$, and since these operators square to zero χ is locally the sum of $\bar{\partial}_A$ and $\bar{\partial}_A^\dagger$-exact pieces. These pieces are easily shown to satisfy the Laplace equation for large r, so the natural *ansatz* is to take them to have the same asymptotics as the L^2 normalizable Green's function for the Laplacian, leading to the behavior

$$\chi \sim \bar{\partial}_A\left(\frac{g^{-1}I^\dagger}{r^2}\right) + \bar{\partial}_A^\dagger\left(g^{-1}\frac{J}{r^2}d\bar{z}_1 \wedge d\bar{z}_2\right),\ r \to \infty. \tag{5.7}$$

Here, I is some $k \times n$ matrix and J is some $n \times k$ matrix, and again there is some natural $U(K)$ action.

In this way we have produced a set of linear operators (B_1, B_2, I, J) between the vector spaces K and N. These may be represented as matrices, up to an action of $U(K)$ reflecting the arbitrariness in the choice of basis of K.

By a tedious Green's function computation (see [6]), one shows that (B_1, B_2, I, J) satisfy the equations, which are invariant under the $U(K)$ action:

$$\begin{aligned} [B_1, B_2] + IJ &= 0 \\ \left[B_1, B_1^\dagger\right] + \left[B_2, B_2^\dagger\right] + II^\dagger - J^\dagger J &= 0. \end{aligned} \tag{5.8}$$

We refer to a multiplet of (B_1, B_2, I, J) satisfying these equations as ADHM data. In this way one has produced a map

$$\text{Instanton gauge field } A \text{ with charge } k \longrightarrow \text{ADHM Data up to } U(K) \text{ action.} \tag{5.9}$$

The claim of the ADHM construction is that this map is an isomorphism, namely the gauge field A may be uniquely reconstructed from the $U(K)$ orbit of ADHM data. The claim is demonstrated by constructing the inverse map explicitly.

Given ADHM data (B_1, B_2, I, J), an instanton solution may be uniquely (re)constructed. We review the main point here, namely how to obtain a rank n vector bundle with anti-self-dual connection A and the matrix χ of normalized solutions to Dirac equation from ADHM data, see [6] for further details.

For each $z = (z_1, z_2) \in \mathbb{C}^2$ define the operator $\mathscr{D}_z^\dagger : (K \otimes \mathbb{C}^2) \oplus N \to K \otimes \mathbb{C}^2$ by

$$\mathscr{D}_z^\dagger := \begin{pmatrix} B_1 - z_1 & B_2 - z_2 & I \\ -B_2^\dagger + \bar{z}_2 & B_1^\dagger - \bar{z}_1 & -J^\dagger \end{pmatrix} \tag{5.10}$$

One verifies using the ADHM equations that the operator $\mathscr{D}_z^\dagger \mathscr{D}_z : K \otimes \mathbb{C}^2 \to K \otimes \mathbb{C}^2$ may be written as $\mathscr{D}_z^\dagger \mathscr{D}_z = \Delta_z \otimes \mathbb{1}_{\mathbb{C}^2}$, with

$$\Delta_z = (B_1 - z_1)(B_1^\dagger - \bar{z}_1) + (B_2 - z_2)(B_2^\dagger - \bar{z}_2) + II^\dagger \in \mathrm{End}(K). \tag{5.11}$$

Note that $\Delta_z \geq 0$, and restricting to the locus of ADHM data such that $\Delta_z > 0$ for all $z \in \mathbb{C}^2$ then $\mathscr{D}_z$ has no kernel and $\mathscr{D}_z^\dagger$ is surjective with $\dim \ker \mathscr{D}_z^\dagger = n$. Then the space of pairs

$$E = \{(z, v) \in \mathbb{C}^2 \times ((K \otimes \mathbb{C}^2) \oplus N) | v \in \ker \mathscr{D}_z^\dagger\} \tag{5.12}$$

defines a rank n vector bundle over $\mathbb{C}^2$. The claim is that E naturally carries a connection A such that $F_A^+ = 0$, and that sections of $S \otimes E$ may be identified with spinors charged in the fundamental of $U(n)$.

To see that E has a natural connection, note that it is a sub-bundle of the trivial bundle over $\mathbb{C}^2$ with fiber $(K \otimes \mathbb{C}^2) \oplus N$. The trivial bundle naturally carries the trivial connection, which is just the d

operator on appropriate vector-valued functions on $\mathbb{C}^2$. If $\pi_z : (K \otimes \mathbb{C}^2) \oplus N \to \ker \mathscr{D}_z^\dagger$ denotes orthogonal projection in the natural hermitian structure, then $D := \pi_z \circ d$ defines a connection on E, and one verifies that its curvature is anti-self-dual.

To make the above remarks more explicit, note that the fiber of E is a vector space of dimension n and may therefore be identified with N. Let $\Psi_z : N \to (K \otimes \mathbb{C}^2) \oplus N$ be the identifying map; thus it is a solution to

$$\mathscr{D}_z^\dagger \Psi_z = 0. \tag{5.13}$$

The columns of Ψ_z define a basis for the fiber of E at each point z–by taking this basis to be orthonormal one may assume $\Psi_z^\dagger \Psi_z = \mathbb{1}_N$. Then the connection A on E defined via the projection above is

$$A = \Psi_z^\dagger d\Psi_z. \tag{5.14}$$

One readily verifies that this is indeed a $U(n)$ connection, and by direct computation demonstrates that $F_A^+ = 0$. Note that A is invariant under the $U(K)$ action on the ADHM data.

The ADHM data also allow one to describe solutions to the Dirac equation valued in E. Write

$$\Psi_z = \begin{pmatrix} v_+ \\ v_- \\ \xi \end{pmatrix} \tag{5.15}$$

for $v_\pm : N \to K$, $\xi : N \to N$, then upon identifying $S_- \cong \Omega^{0,1}$,

$$\chi = v_+^\dagger \Delta_z^{-1} d\bar{z}_1 + v_-^\dagger \Delta_z^{-1} d\bar{z}_2 \tag{5.16}$$

is an $n \times k$ matrix of one-forms (in more intrinsic language, it is a map from K to sections of $\Omega^{0,1} \otimes E$–$\Psi\chi$ takes values in sections of $\Omega^{0,1} \otimes E$) satisfying the Dirac equation $\bar{\partial}_A \chi = \bar{\partial}_A^\dagger \chi = 0$.

This is readily demonstrated by straightforward computation.

This apparatus of constructions allows one to define the moduli space $\mathscr{M}(k,n)$ in purely finite-dimensional terms:

$$\mathscr{M}(k,n) = \left\{ (B_1,B_2,I,J) \in (\mathrm{End}(K)\otimes\mathbb{C}^2)\oplus\mathrm{Hom}(N,K)\oplus\mathrm{Hom}(K,N) \,\middle|\, \text{Equations (5.8) are satisfied and } \mathscr{D}_z \text{ is injective } \forall z\in\mathbb{C}^2 \right\} \Big/ U(K). \tag{5.17}$$

Note, in this description, it is a simple parameter count to see that $\dim\mathscr{M}(k,n) = 4kn$. It is also not difficult to see that $\mathscr{M}(k,n)$ is a smooth manifold.

Compactness, Singularities and Resolution

Using the ADHM construction, it is possible to describe the geometric properties of the instanton moduli space more concretely. The injectivity of $\mathscr{D}_z$ ensures that $\mathscr{M}(k,n)$ is a smooth manifold. However, $\mathscr{M}(k,n)$ is not compact. In order to integrate over it, one must define a suitable compactification.

Note that there are actually two sources of noncompactness for $\mathscr{M}(k,n)$. One is due to the fact that the space of matrices (B_1,B_2,I,J) is non-compact and that the equations defining ADHM data are homogeneous under scaling these variables. This type of non-compactness is an infrared effect, and is due to the fact that instantons can grow arbitrarily large and/or escape to infinity. These effects are a consequence of defining a quantum field theory in infinite volume, and later we explain how to deal with them by Ω-deformation.

The other source of non-compactness is the requirement that $\mathscr{D}_z$ be injective for all $z\in\mathbb{C}^2$–it's possible to find a sequence of (B_1,B_2,I,J) which converges to a configuration such that $\mathscr{D}_z$ has a zero mode for some z. The Donaldson-Uhlenbeck compactification of the moduli space [10] simply adds in the requisite points as limits of these sequences (in other words, it is given by the same definition as above but without the constraint on the rank of $\mathscr{D}_z$). While sufficient for some purposes, this compactification has the undesirable feature that the space is singular at the added points. These singularities are known as "small instanton singularities", because $\mathscr{D}_z$ can develop zero modes at particular values of $z \in \mathbb{C}^2$. These points may be interpreted as locations of instantons that have shrunk to zero size.

The physical reason that these singularities must be dealt with carefully is that this type of non-compactness is intimately related with the ultraviolet cutoff needed to produce a well-defined quantum field theory [28]. It is physically unreasonable to discuss instantons of characteristic size smaller than this cutoff scale. For computations of physically relevant quantities, the precise value of the cutoff should be irrelevant so long as it is much shorter than all other length scales in the theory.

Thus, one would like to arrive at a compactification of the instanton moduli space which has the property that small instantons are forbidden, and instantons are frozen at some cutoff scale. The construction is known in the mathematical literature as the Gieseker-Nakajima compactification [20], [35], and is achieved as follows. Fix a real number $\zeta > 0$, then

$$\begin{aligned} \overline{\mathscr{M}}(k,n) = \Bigg\{ (B_1,B_2,I,J) \in (\mathrm{End}(K)\otimes\mathbb{C}^2) \oplus \mathrm{Hom}(N,K) \oplus \mathrm{Hom}(K,N) \Bigg| \\ [B_1,B_2] + IJ = 0 \\ \left[B_1,B_1^\dagger\right] + \left[B_2,B_2^\dagger\right] + II^\dagger - J^\dagger J = \zeta \mathbb{1}_K \Bigg\} \Bigg/ U(K). \end{aligned} \tag{5.18}$$

One can prove the following key result about this moduli space (see [41] for an argument), known as the stability theorem ($GL(K)$ denotes the group of invertible linear maps from the space K to itself–upon picking a basis it may be identified with $GL(k,\mathbb{C})$):

$$\overline{\mathscr{M}}(k,n) = \left\{ (B_1,B_2,I,J) \in (\mathrm{End}(K)\otimes\mathbb{C}^2)\oplus\mathrm{Hom}(N,K)\oplus\mathrm{Hom}(K,N) \,\middle|\, [B_1,B_2]+IJ=0, \text{ and } \mathbb{C}[B_1,B_2]I(N)=K \right\} \Big/ GL(K). \tag{5.19}$$

The requirement $\mathbb{C}[B_1,B_2]I(N)=K$ is known as the stability condition. From this description, it follows that $\overline{\mathscr{M}}(k,n)$ is a smooth complex manifold. It also admits a map $\overline{\mathscr{M}}(k,n)\to\mathscr{M}(k,n)^{DU}$ which is a resolution of singularities (the superscript DU denotes Donaldson-Uhlenbeck compactification). Physically, this reflects the obvious fact that at length scales $\gg\sqrt{\zeta}$, configurations in $\overline{\mathscr{M}}(k,n)$ look like ordinary instantons, and the configurations toward the small instanton limit are "smoothed out" by the cutoff (which can be interpreted as reflecting ignorance about hidden ultraviolet degrees of freedom). We will not explore the physical interpretation of $\overline{\mathscr{M}}(k,n)$ further, but direct readers to [47], [38] for more information.

As an aside, we note that the stability theorem establishes that $\overline{\mathscr{M}}(k,n)$ is also the moduli space of torsion-free sheaves on $\mathbb{CP}^2$ of rank n and second Chern class k which have a fixed trivialization over the $\mathbb{CP}^1$ at infinity [35]. It can also be thought of as parameterizing instantons in gauge theory defined on "noncommutative $\mathbb{R}^4$" [47].

It is the smooth space $\overline{\mathscr{M}}(k,n)$ that should be used in localization calculations, since it has more desirable properties and reflects a more sensible choice of ultraviolet regularization. It is not a compact space, but provides a partial compactification of the naive instanton moduli space $\mathscr{M}(k,n)$ because it adds in limit configurations for small instantons. It does not cure the problem of large instantons, but this is a separate problem addressed by the Ω-deformation.

Ω-deformation

The final ingredient needed to perform the instanton counting calculations in $\mathcal{N}=2$ gauge theories is known as the Ω-deformation [39], [45]. There are many ways to define it, for simplicity we give only the definition within the framework of the twisted $\mathcal{N}=2$ theories and consider pure $\mathcal{N}=2$ super-Yang-Mills, returning to the issue of matter later on. We replace the integrals over $\mathscr{M}(k,n)$ with integrals over the partial compactification $\overline{\mathscr{M}}(k,n)$.

As pointed out in the last chapter, the scalar supercharge δ of the twisted $\mathcal{N}=2$ theory acts as an equivariant de Rham differential in field space, which is equivariant with respect to the group of gauge transformations. Upon reduction of the path integral to zero modes, one produces an integral in equivariant cohomology of $\overline{\mathscr{M}}(k,n)$ where the Coulomb moduli of the gauge theory become identified with the equivariant parameters associated to the (maximal torus of) the group $U(n)$ of constant gauge transformations acting on $\overline{\mathscr{M}}(k,n)$. In fact, only the group $SU(n)/\mathbb{Z}_n$ acts nontrivially on the moduli space, but we ignore this subtlety. The Ω-deformed gauge theory is defined by working equivariantly also with respect to the (maximal torus of the) $SO(4)$ symmetries of spacetime. Such a rotation is characterized by two parameters $\varepsilon_1, \varepsilon_2$ such that $(z_1, z_2) \mapsto (e^{-i\beta\varepsilon_1}, e^{-i\beta\varepsilon_2} z_2)$. This transformation is generated by some vector field V_ε^μ, and the Ω-deformation corresponds to replacing the scalar supercharge Q discussed previously with the deformed supercharge $Q + V_\varepsilon^\mu G_\mu$, with G_μ defined as in the previous section. The corresponding differential on the fields is denoted δ_ε.

The corresponding supersymmetry variatons are

$$\delta_\varepsilon A = \psi \qquad \delta_\varepsilon \phi = i_{V(\varepsilon_1,\varepsilon_2)}\psi \qquad \delta_\varepsilon \chi^+ = H^+$$
$$\delta_\varepsilon \psi = D_A\phi + i_{V(\varepsilon_1,\varepsilon_2)}F_A \quad \delta_\varepsilon \bar{\phi} = \eta \qquad \delta_\varepsilon H^+ = [\chi^+,\phi] + D_A i_{V(\varepsilon_1,\varepsilon_2)}\chi^+ + i_{V(\varepsilon_1,\varepsilon_2)}D_A\chi^+ \tag{5.20}$$
$$\delta_\varepsilon \eta = [\bar{\phi},\phi] + i_{V(\varepsilon_1,\varepsilon_2)}D_A\bar{\phi}$$

and the corresponding modification of the action to preserve δ_ε symmetry is

$$\begin{aligned} S_\varepsilon =& -\frac{2\pi i\tau}{8\pi^2}\int_{\mathbb{R}^4}\operatorname{tr} F_A\wedge F_A \\ &+ \delta_\varepsilon\left(\int_{\mathbb{R}^4} -\operatorname{tr}\psi\wedge\star(D_A\bar{\phi}+i_{V_\varepsilon}F_A) + \operatorname{tr}\eta\star([\bar{\phi},\phi]+i_{V_\varepsilon}D_A\phi) - \operatorname{tr}\chi^+\left(F_A^+ - \frac{g_{YM}^2}{8}H^+\right)\right). \end{aligned} \tag{5.21}$$

This action defines the Ω-deformed theory.

Because the geometric meaning of the terms is the same, the χ^+ integral once again restricts to $F_A^+ = 0$ and the integral over nonzero modes of $\phi,\bar{\phi}$ together with η implements the division by $\mathscr{G}_\infty$. The only difference is that the vector field on $\overline{\mathscr{M}}(k,n)$ is now the one descending from $D_A\bar{\phi} + i_{V_\varepsilon}F_A$, denoted by $V(\bar{\mathbf{a}},\bar{\varepsilon}_1,\bar{\varepsilon}_2)$. It generates the action of the maximal torus $\mathbb{T}^{n+2}$ of the group of symmetries acting on $\overline{\mathscr{M}}(k,n)$, where $\mathbb{T}^2 \subset SO(4)$. Then the instanton contribution to the partition function of the Ω-deformed theory is

$$Z_{\text{inst}}(\mathbf{a},\varepsilon_1,\varepsilon_2;\mathfrak{q}) = \sum_{k\geq 0}\mathfrak{q}^k\int_{\overline{\mathscr{M}}(k,n)} e^{-\delta_\varepsilon(g(V(\bar{\mathbf{a}},\bar{\varepsilon}_1,\bar{\varepsilon}_2),\psi))}. \tag{5.22}$$

Now, $\mathbb{T}^{n+2}$ acts with a finite number of isolated fixed points on $\overline{\mathscr{M}}(k,n)$, so the integral may easily be done by localization. The fixed point calculation will be discussed shortly. Because

$e^{-\delta_\varepsilon(...)} = 1 + \delta_\varepsilon(...) = 1$ in cohomology, one often writes the instanton integral as simply

$$\int_{\overline{\mathscr{M}}(k,n)} 1 = \sum_{\text{fixed pt.}} \frac{1}{\prod(\text{weights})}. \tag{5.23}$$

In the equality we have used the localization formula. Note that the integral of 1 does not really make sense, but it is understood in the sense of equivariant cohomology–the RHS is essentially the definition of the LHS using localization. We employ this shorthand frequently in what follows. Alternatively, one may think of the operation of integration algebraically as simply a pushforward $H^*_G(\overline{\mathscr{M}}(k,n)) \to H^*_G(\mathrm{pt})$ in G-equivariant cohomology for symmetry group G.

Now that the integral has been reduced to the application of localization formulas, the problem becomes combinatorial. We now turn to the explicit description of the fixed points of the group action and weights in the tangent space of $\overline{\mathscr{M}}(k,n)$.

Fixed Points and Weights

The localization arguments discussed in the previous section almost complete the explicit solution of the Ω-deformed $\mathscr{N} = 2$ theory, but it remains to show that the maximal torus of $U(n) \times SO(4)$ acts with isolated fixed points and calculate the weights in each subspace. In this section, we sketch a combinatorial description of the fixed points and weights using Young diagrams [35], [39].

The (compactified) instanton moduli space is the set of (B_1, B_2, I, J) satisfying $\mu_{\mathbb{C}}(B_1, B_2, I, J) := [B_1, B_2] + IJ = 0$ (and the stability condition) up to the action of $GL(K)$, where $(B_1, B_2, I, J) \sim (gB_1g^{-1}, gB_2g^{-1}, gI, Jg^{-1})$, with $g \in \mathrm{End}(K)$. Since $\mu_{\mathbb{C}} : (\mathrm{End}(K) \otimes \mathbb{C}^2) \oplus \mathrm{Hom}(N,K) \oplus \mathrm{Hom}(K,N) \to \mathrm{End}(K)$, for each (B_1, B_2, I, J) representing a point in the moduli space there is a complex

$$0 \to \mathrm{End}(K) \to (\mathrm{End}(K) \otimes \mathbb{C}^2) \oplus \mathrm{Hom}(N,K) \oplus \mathrm{Hom}(K,N) \xrightarrow{d\mu_{\mathbb{C}}} \mathrm{End}(K) \to 0. \tag{5.24}$$

The first map sends $\xi \mapsto ([\xi, B_1], [\xi, B_2], \xi I, -J\xi)$. The second map sends $(\delta B_1, \delta B_2, \delta I, \delta J) \mapsto [\delta B_1, B_2] + [B_1, \delta B_2] + \delta IJ + I\delta J = d\mu_{\mathbb{C}}(\delta B_1, \delta B_2, \delta I, \delta J)$. This sequence is in fact a complex because the infinitesimal $GL(K)$ action is a symmetry of $\mu_{\mathbb{C}} = 0$. One proves using the stability condition that this complex has vanishing cohomology in degree zero and degree two, which establishes that $\overline{\mathscr{M}}(k,n)$ is a smooth complex manifold. The holomorphic tangent space of $\overline{\mathscr{M}}(k,n)$ at the point corresponding to (B_1, B_2, I, J) may be identified with H^1 of this complex. For this reason the above complex is known as the tangent complex.

This identifies $T_p\overline{\mathscr{M}}(k,n)$ at any point p as a vector space. However, to compute the weights, one is interested in $T_p\overline{\mathscr{M}}(k,n)$ as a $\mathbb{T}^n \times \mathbb{T}^2$-module, for p a fixed point of $\mathbb{T}^{n+2}$. The advantage of the description using cohomology is that its algebraic nature readily yields representation-theoretic data. But first, one must classify the fixed points p.

The $\mathbb{T}^{n+2}$ action on the space $(\mathrm{End}(K) \otimes \mathbb{C}^2) \oplus \mathrm{Hom}(N,K) \oplus \mathrm{Hom}(K,N)$ is given by, for $(e^{\beta \mathbf{a}}, e^{\beta\varepsilon_1}, e^{\beta\varepsilon_2}) \in \mathbb{T}^{n+2}$:

$$(e^{\beta \mathbf{a}}, e^{\beta\varepsilon_1}, e^{\beta\varepsilon_2}) \cdot (B_1, B_2, I, J) = (e^{-\beta\varepsilon_1} B_1, e^{-\beta\varepsilon_2} B_2, I e^{-\beta \mathbf{a}}, e^{-\beta(\varepsilon_1+\varepsilon_2)} e^{\beta \mathbf{a}} J). \tag{5.25}$$

This is the canonical action induced naturally from the definitions of these operators. Note that the action of $U(n)$ is defined simply by the fact that N splits into one-dimensional eigenspaces N_α each carrying an eigenvalue $e^{\beta a_\alpha}$–we have used the notation $\mathbf{a}$ to denote also the diagonal matrix $\mathrm{diag}(a_1, ..., a_n)$. The geometric construction of ADHM data, where N is identified with the fiber of the bundle E in which the Dirac fermions in the fundamental take values, makes the representation-theoretic content of N evident. Strictly speaking, to get an action of the compact group $\mathbb{T}^{n+2}$ the (complex) variables $\beta, \varepsilon_1, \varepsilon_2, \mathbf{a}$ must be taken to obey some reality conditions; we leave it to the reader to fill in these fairly straightforward details.

This group action descends to $\overline{\mathscr{M}}(k,n)$. For the matrices (B_1, B_2, I, J) to represent a fixed point in the moduli space, the action of $\mathbb{T}^{n+2}$ on them may be undone by a $GL(K)$ transformation. That is, for a given fixed point and generic values of $\varepsilon_{1,2}$ and $\mathbf{a}$ parameters there exists a $\xi \in \mathrm{End}(K)$ such that

$$
\begin{aligned}
[\xi, B_1] &= \varepsilon_1 B_1 \\
[\xi, B_2] &= \varepsilon_2 B_2 \\
\xi I &= I\mathbf{a} \\
J\xi &= (\mathbf{a} - \varepsilon_1 - \varepsilon_2)J.
\end{aligned}
\tag{5.26}
$$

Strictly speaking, one should write $\xi(\mathbf{a}, \varepsilon_1, \varepsilon_2)$ as a linear function of the equivariant parameters, but since we work at fixed $\mathbf{a}, \varepsilon_1, \varepsilon_2$ it is immaterial. The fixed points of $\mathbb{T}^{n+2}$ acting in the moduli space are in one-to-one correspondence with isomorphism classes of representations of the above algebra. These admit the following description.

First, one proves using the stability condition that for any solution to (5.26), $J = 0$. Then one has $[B_1, B_2] = 0$. The space K splits as $K = \oplus_{\alpha=1}^{n} K_\alpha$, where $K_\alpha = \mathbb{C}[B_1, B_2]I(N_\alpha)$. The eigenvalues of ξ on the space K_α are of the form $a_\alpha + (i-1)\varepsilon_1 + (j-1)\varepsilon_2$ for $(i, j) \in \mathbb{Z}_+^2$. In each subspace K_α, starting from the vectors $I(N_\alpha)$ the B's act as “raising operators" for ξ which shift its eigenvalues by $\varepsilon_{1,2}$, and for an appropriate reality condition and generic parameters $\mathbf{a}$, $\varepsilon_{1,2}$ each eigenspace of ξ is orthogonal. By finite-dimensionality, $B_1^i B_2^j I(N_\alpha) = 0$ for i, j large enough, and it is easy to convince oneself that if one arranges the points labeled by (i, j) in a coordinate grid then the statement that K_α is an orthogonal sum of one-dimensional subspaces $B_1^i B_2^j I(N_\alpha)$ allows one to interpret each one-dimensional subspace as a box in a Young diagram $\lambda^{(\alpha)}$. The number of boxes in the diagram $\lambda^{(\alpha)}$ is denoted by $|\lambda^{(\alpha)}|$.

Thus, the isomorphism classes of representations of the algebra (5.26) are in one-to-one correspondence with n-tuples of Young diagrams $\underline{\lambda} = (\lambda^{(1)}, ..., \lambda^{(n)})$ such that

$$\sum_{\alpha} |\lambda^{(\alpha)}| = k. \tag{5.27}$$

Therefore, fixed points of the $\mathbb{T}^{n+2}$ action on $\overline{\mathscr{M}}(k,n)$ are labeled by n-tuples of Young diagrams $\underline{\lambda}$ with k boxes in total.

Decomposing K into eigenspaces of the operator $e^{\beta\xi(\mathbf{a},\varepsilon_1,\varepsilon_2)}$ gives its content as a $\mathbb{T}^{n+2}$ representation at a fixed point. In what follows, it will be useful to identify vector spaces with their $\mathbb{T}^{n+2}$ characters (really, one should consider these "spaces" as $\mathbb{T}^{n+2}$ equivariant vector bundles over the compactified instanton moduli space and consider their classes in equivariant K-theory), so that

$$\begin{aligned} N &= \sum_{\alpha=1}^{n} e^{\beta a_\alpha} \\ K &= \sum_{\alpha=1}^{n} e^{\beta a_\alpha} \sum_{(i,j)\in\lambda^{(\alpha)}} q_1^{i-1} q_2^{j-1}. \end{aligned}$$

We introduce the notation $q_1 = e^{\beta\varepsilon_1}$ and $q_2 = e^{\beta\varepsilon_2}$.

One verifies that for (B_1, B_2, I, J) representing a fixed point of the $\mathbb{T}^{n+2}$ action, the following complex (which is the same as the tangent complex twisted by one-dimensional representations of $\mathbb{T}^2$ in appropriate places) is $\mathbb{T}^{n+2}$ equivariant with respect to the action (5.25):

$$0 \to q_1 q_2 \mathrm{End}(K) \to q_2 \mathrm{End}(K) \oplus q_1 \mathrm{End}(K) \oplus q_1 q_2 \mathrm{Hom}(N,K) \oplus \mathrm{Hom}(K,N) \to \mathrm{End}(K) \to 0. \tag{5.28}$$

The maps are the same as for the ordinary tangent complex. Because this complex is $\mathbb{T}^{n+2}$ equivariant, its cohomology furnishes representations of $\mathbb{T}^{n+2}$, and in particular the weights of H^1 correspond to the weights of the $\mathbb{T}^{n+2}$ action on $\overline{\mathscr{M}}(k,n)$ in the holomorphic tangent space at the fixed point $\underline{\lambda}$.

In fact, this setup is typical of a supersymmetric quantum mechanics problem (see Chapter 10 of [23])–one has a space of supersymmetric ground states given by H^* of this complex and wants to understand its decomposition under the relevant symmetry algebra. One may think of the (equivariant) tangent complex as a finite-dimensional version of the Hilbert space of a supersymmetric quantum-mechanical system, and the maps as finite-dimensional versions of supercharges. Then by standard deformation invariance arguments for the (equivariant) Witten index [23], the virtual character of the complex agrees with the alternating sum of characters of the cohomology groups– this is just the statement that $\mathrm{Tr}(-1)^F g$ receives contributions only from supersymmetric ground states.

Together with the vanishing theorem $H^0 = H^2 = 0$, this gives the character of the tangent space to the fixed point $\underline{\lambda}$ as

$$T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n) = NK^* + q_1 q_2 KN^* - (1-q_1)(1-q_2)KK^*. \tag{5.29}$$

Given the above formulas for the characters of N and K, it is straightforward to write this character in terms of combinatorial data associated to the Young diagrams $\underline{\lambda}$ as [35]

$$T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n) = \sum_{\alpha,\beta=1}^{n} e^{a_\alpha - a_\beta}\left(\sum_{\square\in\lambda^{(\alpha)}} q_1^{\ell_\alpha(\square)+1} q_2^{-A_\beta(\square)} + \sum_{\square\in\lambda^{(\beta)}} q_1^{-\ell_\beta(\square)} q_2^{A_\alpha(\square)+1}\right). \tag{5.30}$$

The quantity $\ell_\alpha(\square)$ is known as the leg length of the box $\square$ in the diagram $\lambda^{(\alpha)}$, and likewise $A_\alpha(\square)$ is the arm length. Associated to a partition λ, regarded as a nonincreasing sequence of integers λ_i, one defines the conjugate partition by $\lambda_i^t = \#\{i|\lambda_j \geq i, \text{some } j\}$. Then for $\square$ at position (i,j), $\ell_\alpha(\square) = \lambda_j^{t(\alpha)} - i$ and $A_\beta(\square) = \lambda_i^{(\beta)} - j$. The geometric interpretation of this is given in Figure 5.1.

In summary, the fixed points of the $\mathbb{T}^{n+2}$ action on $\overline{\mathscr{M}}(k,n)$ are classified in terms of n-tuples

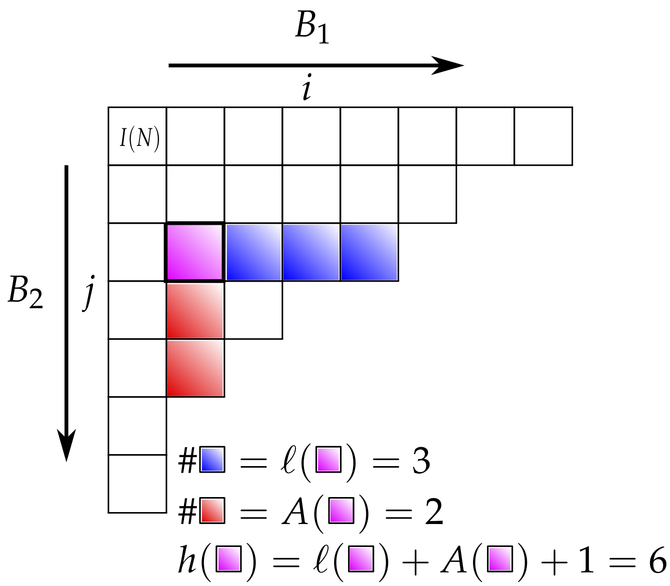

Figure 5.1: In the case $n = 1$, we drop the α label because it is redundant and fixed points are labeled by a single Young diagram such as the one above, corresponding to the partition $\lambda = (7,5,4,3,3,2,1,1)$, with conjugate $\lambda^t = (8,6,5,3,2,1,1)$. Each box in the diagram corresponds to a one-dimensional subspace of K, and K splits as an orthogonal sum over these subspaces–the top left box corresponds to $I(N)$ and a box in position (i,j) corresponds to $B_1^{i-1}B_2^{j-1}I(N)$. The leg and arm length are computed as shown above. Also important is the hook length $h(\square) = \ell(\square) + A(\square) + 1$, which will return later.

of Young diagrams $\underline{\lambda}$. They are isolated and have weights given by some combinatorial formula in terms of data associated to the Young diagrams. From this data, it is possible to compute the partition function of the Ω-deformed theory as a sum over Young diagrams.

Nekrasov Partition Function and Seiberg-Witten Curve

Finally, with all the preliminaries in place we may explain one of the main results, which is the partition function of the Ω-deformed $\mathcal{N} = 2$ theory. This is known also as the Nekrasov partition function and was introduced in [39]. In discussions of Nekrasov partition functions, it is useful to introduce the following notation known as plethystic exponential. Given a virtual character which contains terms e^{A_i} with positive coefficients and terms e^{B_j} with negative coefficients, we define:

$$\mathbb{E}\left[\sum_i e^{A_i} - \sum_j e^{B_j}\right] := \frac{\prod_j B_j}{\prod_i A_i}. \tag{5.31}$$

Note this essentially computes the “supersymmetric path integral" associated to a virtual representation– the “fermionic states" with a minus sign contribute a determinant in the numerator, and the “bosonic states" contribute a determinant in the denominator. The Nekrasov partition function for the pure $\mathcal{N} = 2$ theory under consideration thus far may be written in this notation as (we replace $\mathfrak{q}$ with the dynamical scale Λ that appears since this theory is asymptotically free)

$$Z(\mathbf{a}, \varepsilon_1, \varepsilon_2; \Lambda) = \sum_{k \geq 0} \Lambda^{2nk} \sum_{\substack{\underline{\lambda} \\ |\underline{\lambda}|=k}} \mathbb{E}[T_{\underline{\lambda}} \overline{\mathscr{M}}(k,n)]. \tag{5.32}$$

The limit $\varepsilon_{1,2} \to 0$ of Z recovers the famous Seiberg-Witten solution. In fact, this was Nekrasov's original goal of introducing the Ω-deformation and instanton calculus in this setup: to provide a first-principles derivation of the Seiberg-Witten solution of gauge theory [39]. The main relation is

$$\mathscr{F}_0(\mathbf{a}) = \lim_{\varepsilon_{1,2} \to 0} -\varepsilon_1 \varepsilon_2 \log Z \tag{5.33}$$

where $\mathscr{F}_0$ is a quantity known as the Seiberg-Witten prepotential [51]. For mathematicians, this is actually the definition of the prepotential. More about this will be discussed below. There are several ways of understanding the relationship between Z and Seiberg-Witten theory. One way is to proceed by analogy with matrix models.

The $n = 1$ Case in Detail

The case of $n = 1$, where the gauge group is $U(1)$, is of course somewhat trivial given that the underlying field theory is actually free, but the mathematical structure that arises is nonetheless rich and provides a hint for the case of general n.

The fact that we still have interesting mathematical structures reflects the fact that the partially compactified space $\overline{\mathscr{M}}(k,1)$ is interesting and nontrivial–it is an object known in algebraic geometry as the Hilbert scheme of k points in $\mathbb{C}^2$, $\mathrm{Hilb}_k(\mathbb{C}^2)$.

It is useful to consider a special Ω-background, known as the self-dual Ω-background, for which $\varepsilon_1 + \varepsilon_2 = 0$, so that $\varepsilon_1 = -\varepsilon_2 := g_s$. With this notation, one has

$$\mathbb{E}[T_\lambda \mathrm{Hilb}_k(\mathbb{C}^2)] = g_s^{-2k} \prod_{\square \in \lambda} \frac{1}{h(\square)^2}. \tag{5.34}$$

The quantity $h(\square) = \ell(\square) + A(\square) + 1$ is known as the *hook length* of the box $\square$ in the diagram λ. By the so-called hook length formula [17],

$$\prod_{\square \in \lambda} h(\square) = \frac{|\lambda|!}{\dim \lambda} \tag{5.35}$$

where $\dim \lambda$ is the same number introduced earlier in the present thesis. Thus, the Nekrasov partition function in this case is given simply in terms of the Plancherel measure on partitions:

$$Z = \sum_{k \geq 0} \frac{1}{k!} \left(\frac{\Lambda^2}{g_s^2} \right)^k \sum_{\substack{\lambda \\ |\lambda| = k}} \mathscr{P}(\lambda) = e^{\frac{\Lambda^2}{g_s^2}}. \tag{5.36}$$

As reviewed in chapter 2, this partition function may be expressed as a matrix model which captures the “limit shape" phenomenon of the random partitions. That is, in the limit $g_s \to 0$ the sum over partitions is dominated by a most probable partition of large size described by the Logan-Schepp-Vershik-Kerov profile [27], [59].

The point of this is that it is a generic feature of the Nekrasov partition function–in the limit $\varepsilon_{1,2} \to 0$ the sum over partitions is dominated by a limit shape. The "spectral curve" capturing the limit shape in the matrix model analogy is known as the Seiberg-Witten curve of the gauge theory–in the next section it will be explained how to obtain this curve without using matrix model techniques directly (although using techniques inspired by those in matrix models).

The Y-observables

The combinatorial sum over partitions to which the instanton partition function reduces is reminiscent of the kind of combinatorial object that is the matrix model partition function. It is therefore natural to use similar techniques as in matrix models to understand more about the structure of the instanton partition function.

Recall that the main tool that was useful in solving the matrix model was the construction of the resolvent $\mathrm{tr}(x-\Phi)^{-1}$, as its singularity structure in x effectively encoded the loop equations for the matrix model, which can be interpreted as Dyson-Schwinger equations for the gauge theory. It is therefore natural to look for observables $Y(x)$ in the $\mathcal{N}=2$ theory that depend on some auxiliary spectral parameter $x \in \mathbb{C}$, that have similarly desirable analytic properties.

A natural candidate for such an observable is the characteristic polynomial of the Higgs field $\phi(z,\bar{z})$, where $z=(z_1,z_2) \in \mathbb{C}^2$. One would like to study

$$Y(x) \sim \det(x-\phi(0)). \tag{5.37}$$

Note that the Higgs field must be placed at the origin in order for this observable to preserve the supersymmetry of the Ω-deformed theory. The expression on the RHS is not well-defined because of contact term ambiguities–the operator products of $\phi(z,\bar{z})$ at different spacetime points are singular as the points approach each other. These can be interpreted as quantum effects due to the instanton background.

To find the appropriate definition of the Y-observable that automatically takes into account the non-perturbative corrections/quantum effects, it is best to interpret it as a geometric object associated to $\overline{\mathscr{M}}(k,n)$.

At a point of $\overline{\mathscr{M}}(k,n)$ represented by ADHM data (B_1,B_2,I,J) there is a complex of vector spaces, known as the tautological complex

$$0 \to K \to (K \otimes \mathbb{C}^2) \oplus N \to K \to 0. \tag{5.38}$$

The map first map acts by

$$v \mapsto \begin{pmatrix} -B_1 v \\ B_2 v \\ J v \end{pmatrix} \tag{5.39}$$

for $v \in K$. The second map acts by

$$\begin{pmatrix} v_+ \\ v_- \\ \xi \end{pmatrix} \mapsto B_1 v_- + B_2 v_+ + I\xi \tag{5.40}$$

for $v_\pm \in K$ and $\xi \in N$. Allowing this to vary across all points in $\overline{\mathscr{M}}(k,n)$ gives a complex of (equivariant) vector bundles, and by abuse of notation we also denote these bundles by K and N.

After a twist by representations of $\mathbb{T}^2 \subset SO(4)$ this complex becomes $\mathbb{T}^{n+2}$-equivariant:

$$0 \to q_1 q_2 K \to q_2 K \oplus q_1 K \oplus N \to K \to 0. \tag{5.41}$$

Define the object S to be the virtual equivariant Chern character of this complex:

$$S = \mathrm{ch}(K) + \mathrm{ch}(K \otimes \Lambda^2 \mathbb{C}^2) - \mathrm{ch}(N) - \mathrm{ch}(K \otimes \mathbb{C}^2). \tag{5.42}$$

This is understood in the usual sense of equivariant Chern character–the object ch is a formal character valued in $H^*_{\mathbb{T}^{n+2}}(\overline{\mathscr{M}}(k,n))$ using the Chern roots of the various bundles together with their weights as torus representations. The degree zero piece at a fixed point of the $\mathbb{T}^{n+2}$ action reduces to the ordinary virtual character of the fiber. The Y-observable, as an element of the equivariant cohomology ring of $\overline{\mathscr{M}}(k,n)$, is defined to be the plethystic exponential of the virtual character of the tautological complex: (c_x denotes Chern polynomial, in other words the equivariant top Chern class with respect to the $\mathbb{C}^\times$ symmetry scaling the fiber of the bundle, with equivariant parameter x):

$$Y(x) = \mathbb{E}[e^x S^*] = \frac{c_x(N^*) c_x(K^* \otimes \mathbb{C}^{2*})}{c_x(K^*) c_x(K^* \otimes (\Lambda^2 \mathbb{C}^2)^*)}. \tag{5.43}$$

At a fixed point, this may be evaluated using the known weights of the representations (we set $\beta = 1$ for simplicity, it may be always reinstated by dimensional analysis):

$$Y(x)\Big|_{\underline{\lambda}} = \prod_{\alpha=1}^{n} (x - a_\alpha) \prod_{\square \in \lambda^{(\alpha)}} \frac{(x - a_\alpha - c(\square) - \varepsilon_1)(x - a_\alpha - c(\square) - \varepsilon_2)}{(x - a_\alpha - c(\square))(x - a_\alpha - c(\square) - \varepsilon_1 - \varepsilon_2)}. \tag{5.44}$$

The notation $c(\square)$ has been introduced: for a box in the Young diagram $\lambda^{(\alpha)}$ at position (i,j), it is defined to be $c(\square) = (i-1)\varepsilon_1 + (j-1)\varepsilon_2$–this is referred to as the content of the box $\square$.

In the limit where ε's vanish, this reduces to the characteristic polynomial evaluated at the vacuum expectation value of ϕ. The full definition of Y may be thought of as including quantum corrections, with the inclusion of the vector spaces K in the tautological complex signifying the presence of fermionic zero modes in the instanton background of charge k that lead to the need to correct the naive characteristic polynomial. In particular, the corrected version of $Y(x)$ is no longer a polynomial but is a rational function.

The main interesting feature of Y is that there are several cancellations occurring in the ratio defining Y above, so that it may be rewritten as follows. Let $\partial_+\lambda$ be the outer boundary of the Young diagram λ, defined to be the set of boxes which can be added to the Young diagram while keeping it a Young diagram. Likewise, the inner boundary $\partial_-\lambda$ is the set of boxes which may be removed from the Young diagram while keeping it a Young diagram (an example is shown in Figure 5.2). It is straightforward to show that Y may be written as

$$Y(x)\Big|_{\underline{\lambda}} = \prod_{\alpha=1}^{n} \frac{\prod_{\square\in\partial_+\lambda^{(\alpha)}}(x-a_\alpha-c(\square))}{\prod_{\square\in\partial_-\lambda^{(\alpha)}}(x-a_\alpha-c(\square)-\varepsilon_1-\varepsilon_2)}. \tag{5.45}$$

This allows for further understanding of Y: its analytic properties reflect the possibility of adding or removing a single instanton from the system. This is again analogous to the resolvent in matrix models, the poles of which encoded information about the spectrum of the random matrix. Therefore, Ward identities which constrain the analytic behavior of Y may be understood as organizing information about the behavior of the field theory under adding or removing instantons. Since this corresponds to changing the topological sector of field space, this may be understood as describing a symmetry of the path integral of the theory under a “large" (not continuously connected to the identity) deformation of the integration contour.

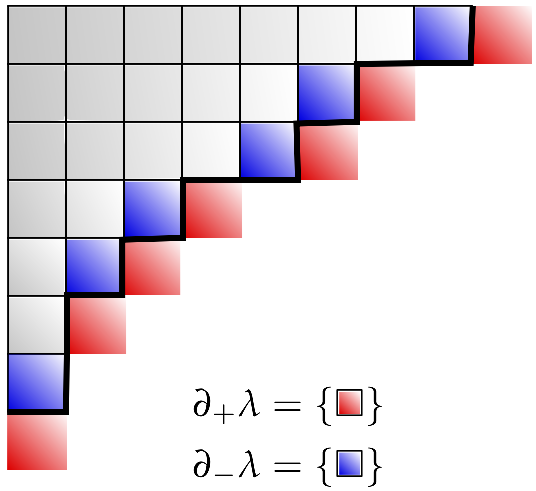

Figure 5.2: For the partition $\lambda = (7,5,4,3,2,1,1)$, the blue boxes are those in $\partial_-\lambda$, and the red are those in $\partial_+\lambda$. Importantly, those in $\partial_+\lambda$ do not belong to the original Young diagram, while those in $\partial_-\lambda$ do.

We employ the notation

$$\langle \mathscr{O} \rangle = \frac{1}{Z} \sum_{k \geq 0} \Lambda^{2nk} \sum_{\substack{\underline{\lambda} \\ |\underline{\lambda}|=k}} \mathbb{E}[T_{\underline{\lambda}} \overline{\mathscr{M}}(k,n)] \mathscr{O} \Big|_{\underline{\lambda}} \tag{5.46}$$

for a normalized expectation value of an observable.

In the pure $\mathscr{N} = 2$ Yang-Mills theory with gauge group $U(n)$, one can prove directly that

$$\left\langle Y(x+\varepsilon_1+\varepsilon_2) + \frac{\Lambda^{2n}}{Y(x)} \right\rangle \tag{5.47}$$

has no poles in x by noting that those from the first and second terms cancel each other. As mentioned above, the statement that this object has no poles reflects a transformation property of the path integral under a “large" deformation of the contour so the above equation is referred to as a nonperturbative Dyson-Schwinger equation.

The nonperturbative Dyson-Schwinger equation immediately leads to a description of the Seiberg-Witten geometry of the gauge theory. In the limit $\varepsilon_{1,2} \to 0$, as discussed above the sum over partitions localizes on a saddle point configuration known as the limit shape. We let $z = \lim_{\varepsilon_{1,2}\to 0}\langle Y(x)\rangle$ be the value of Y evaluated on the limit shape. Since $Y(x)$ is a degree n monic polynomial for large x, the nonperturbative Dyson-Schwinger equation in the limit $\varepsilon_{1,2} \to 0$ leads to

$$z + \frac{\Lambda^{2n}}{z} = P(x) \tag{5.48}$$

where $P(x) = x^n + \sum_{i=0}^{n-1} u_i x^{n-i}$, for some constants u_i. In the classical limit these constants are elementary symmetric functions of the a_α, but they recieve quantum corrections from instantons. This equation describes a hyperelliptic curve of genus $g = n-1$ known as the Seiberg-Witten curve, and one verifies (by making similar arguments as in the case of the matrix model, for a similar choice of A and B cycles) that

$$\begin{aligned} a_\alpha &= \frac{1}{2\pi i}\oint_{A^\alpha} x\frac{dz}{z} \\ \frac{\partial \mathscr{F}_0}{\partial a^\alpha} &= \oint_{B_\alpha} x\frac{dz}{z}. \end{aligned} \tag{5.49}$$

These are the same special geometry relations encountered previously, and determine the prepotential $\mathscr{F}_0$ in terms of the underlying algebraic curve. In the physical interpretation of $\mathscr{N} = 2$ theories, the prepotential essentially uniquely characterizes the low-energy effective action of the theory to lowest order in a derivative expansion. This solution of $\mathscr{N} = 2$ super-Yang-Mills (the determination of the exact low-energy effective action to lowest order in derivative expansion) is a famous result known as the Seiberg-Witten solution of gauge theory. The Ω-deformation and instanton counting method, while requiring lots of background, gives a direct first-principles derivation of this result (originally, Seiberg and Witten guessed the form of the effective action using ingenious symmetry arguments).

Aside: Physics of Seiberg-Witten Solution

In order to keep the progression and presentation reasonably brief and self-contained, we have presented Seiberg-Witten theory only as some limit of nonperturbative Dyson-Schwinger equations, by analogy to the behavior of matrix models in the planar limit. However, it is of great interest in physics because it captures the low energy behavior of $\mathscr{N}=2$ theories, and provided some of the first evidence for dualities that led to the insights on dualities in string theory. We briefly sketch the physical interpretation of these results here.

Four-dimensional $\mathscr{N}=1$ theories have a well-known superspace formalism, and there is an analogous superspace formalism for four-dimensional $\mathscr{N}=2$ theories, albeit the details are more involved. The $\mathscr{N}=2$ chiral superfield Ψ organizes the vector multiplet $(\Phi, \lambda^{\alpha I}, A)$ and since the number of supercharges are doubled, the chiral part of superspace now consists of four fermionic directions. The full superspace measure would be $d^4 x d^8 \theta$, but the action of $\mathscr{N}=2$ theories consists only of an integral over the chiral piece $d^4 x d^4 \theta$: it is given by

$$S \propto \operatorname{Im} \int d^4 x d^4 \theta \mathscr{F}_0(\Psi). \tag{5.50}$$

Ψ is taken to be in the adjoint representation of the gauge group and the usual $\mathscr{N}=2$ super-Yang-Mills action is recovered by choosing $\mathscr{F}_0(\Psi) \propto \operatorname{tr} \Psi^2$. Vacua of the theory are minima of the bosonic potential which is essentially $\left\| [\phi, \phi^\dagger] \right\|^2$ as discussed before. The Coulomb parameters $\mathbf{a}$ are coordinates on the moduli space of vacua, and their vacuum expectation values give masses to some of the other fields. By the supersymmetric Higgs mechanism, for generic choice of $\mathbf{a}$ the gauge group is broken $U(n) \to U(1)^n$ at low energies, the massive fields may be integrated out and the Wilsonian effective action is now of type (5.50) but with $\mathscr{F}_0(\mathscr{A}(x))$ a general function of the low-energy (super)fields, which in general receives quantum corrections from integrating out the heavy fields and fast modes.

Supersymmetry restricts the perturbative corrections to $\mathscr{F}_0$ to be one-loop exact, so the only possibility which remains is that of instanton corrections, which themselves are severely constrained by holomorphy of $\mathscr{F}_0$. In their original paper, Seiberg and Witten determined $\mathscr{F}_0$ by considering the BPS spectrum of the $SU(2)$, $\mathscr{N}=2$ theory, making a certain physically well-motivated *ansatz* and imposing consistency conditions [51]. This implicitly determined the form of the instanton corrections, but it was out of reach of analytic techniques at the time to obtain such corrections by direct computation. Instanton counting was introduced as a framework to compute these corrections directly, without making any *ansatz* or assumptions.

The form of the effective action explains the relation between the partition function and prepotential– in general, for any statistical mechanical system in some volume V one has $Z = \exp\{-VF\}$ where F is the free energy (density), which in the large volume limit is just the ground state energy (density).

In the case of an $\mathscr{N}=2$ theory in the Ω-background, the superspace integral $d^4xd^4\theta$ should be interpreted as integration in equivariant cohomology of $\mathbb{R}^4$ for the the differential $d+i_V$ (the supercharge preserved by Ω-background is just the lift of this differential to field space, also twisted by global gauge transformations), and neglecting the variation of $\mathbf{a}(x)$ in space from its vacuum expectation value (valid in the large volume limit) the localization formula gives (by supersymmetry, the antiholomorphic piece is δ-exact and decouples)

$$\int d^4xd^4\theta\mathscr{F}_0(\mathscr{A}(x)) \to \frac{1}{\varepsilon_1\varepsilon_2}\mathscr{F}_0(\mathbf{a}). \tag{5.51}$$

Since the equivariant volume of $\mathbb{R}^4$ (that is, the equivariant integral of 1) is just $1/\varepsilon_1\varepsilon_2$ (there is only one fixed point and these are its weights), the large volume limit is $\varepsilon_1, \varepsilon_2 \to 0$ so this is how one extracts the prepotential.

From knowledge of the Seiberg-Witten prepotential, the low-energy behavior of the theory is completely determined and in principle any question may be answered by an appropriate analysis of the effective action. We do not go further into the low energy physics of $\mathscr{N} = 2$ theories (which is itself fascinating, with connections to the so-called wall-crossing phenomenon, Hitchin systems, and string theory [18], [19]) because it is well-documented in other places (see for example [54] and references therein).

Branes and Generalized ADHM Data for Quiver Gauge Theories

It is remarkable that the ADHM construction of instantons can be obtained via string theory [11], [62]. The key point is that the ADHM description of the moduli space $\mathscr{M}(k,n)$ is a hyper-Kähler quotient, which in turn describes the Higgs branch of some $U(k)$ gauge theory. The gauge theory in question can be interpreted as living on the worldvolume of the instantons themselves, regarded as a collection of k $D(-1)$-branes bound to a stack of n $D3$ branes in Type IIB string theory (see [55] for a review of this point of view). Passing to the partially compactified instanton moduli space corresponds to turning on the B-field in string theory [47].

The gauge theory living on the worldvolume of a stack of $D3$ branes in flat space is $\mathscr{N} = 4$ supersymmetric Yang-Mills theory, while the above constructions computed the partition function for $\mathscr{N} = 2$ supersymmetric Yang-Mills. A consequence of [41] is that it is possible to obtain the instanton contributions of the $\mathscr{N} = 4$ theory from the point of view of the $D(-1)$ branes, by regarding instanton counting in the $\mathscr{N} = 4$ theory as the problem of enumerating supersymmetric bound states of the $D(-1)-D3$ system in the Ω-background. The main result is that in addition to the ADHM data (B_1,B_2,I,J), there are two more $k\times k$ matrix fields (B_3,B_4), which modify the equations for vacua as

$$\begin{aligned}
[B_1,B_2]+IJ+[B_3,B_4]^\dagger &= 0\\
\left[B_1,B_1^\dagger\right]+\left[B_2,B_2^\dagger\right]+\left[B_3,B_3^\dagger\right]+\left[B_4,B_4^\dagger\right]+II^\dagger-J^\dagger J &= \zeta\mathbb{1}_K\\
[B_1,B_4]+[B_2,B_3]^\dagger &= 0\\
[B_1,B_3]+[B_4,B_2]^\dagger &= 0\\
B_3I+B_4^\dagger J^\dagger &= 0\\
B_4I-B_3^\dagger J^\dagger &= 0.
\end{aligned}\tag{5.52}$$

The equation $\left[B_1,B_1^\dagger\right]+\ldots$ may be replaced with a stability condition, and by adding together the norm squares of the remaining equations, one notes that the cross terms cancel so that the space of solutions to the above equations modulo $U(K)$ is equivalent to the space of the solutions to the holomorphic equations

$$\begin{aligned}
[B_1,B_3]=[B_1,B_4]=[B_2,B_3]=[B_2,B_4] &= 0\\
[B_1,B_2]+IJ=[B_3,B_4] &= 0\\
B_3I=B_4I &= 0\\
JB_3=JB_4 &= 0\\
\mathbb{C}[B_1,B_2,B_3,B_4]I(N) &= K
\end{aligned}\tag{5.53}$$

modulo $GL(K)$. It is clear that the stability condition, together with the above equations, implies $B_3=B_4=0$ on any solution so that the space of solutions is once again the instanton moduli space $\overline{\mathscr{M}}(k,n)$. However, the instanton partition function is different in this case because the deformation complex for the above equations is modified by the presence of the B_3,B_4 fields, in spite of the fact that they vanish on the solution set. Mathematically, the modified equations describe the same underlying manifold $\overline{\mathscr{M}}(k,n)$, but define a different virtual fundamental class, in this case the Euler class of an obstruction bundle. Physically, in the problem of enumeration of the $D(-1)-D3$ bound states, the partition function may be reduced to a kind of supersymmetric matrix model involving the B's, and in this case there are additional fermion zero modes in the matrix model.

The supersymmetric partition function of the system can be computed by considering the deformation complex for the modified system of equations. In the spirit of Nekrasov's calculus, it is desirable to work equivariantly with respect to all symmetries of the problem. The modified ADHM equations admit an additional equivariant parameter, which may eventually be identified with the mass of an adjoint hypermultiplet field. The virtual character of the deformation complex defines a class in the equivariant K-theory of the moduli space of solutions (which is in this case once again the compactified instanton moduli space) known as the virtual tangent space.

Consider multiplying each B_α by a phase factor, $B_\alpha \to q_\alpha B_\alpha$. Scaling $J \to q_1 q_2 J$, it is evident that this is a symmetry of the equations provided $q_1 q_2 q_3 q_4 = 1$–in terms of equivariant parameters for the rotations, one must have $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4 = 0$. This equation means that there is a $\mathbb{T}^{n+3}$ worth of symmetries acting on these equations.

The tangent space to the moduli space can be modeled, as a real vector space, as H^1 of the following complex:

$$
\begin{aligned}
&0 \to \mathrm{Lie}U(K) \to \mathrm{Hom}(K, K \otimes \mathbb{C}^2_{12}) \oplus \mathrm{Hom}(N,K) \oplus \mathrm{Hom}(K,N) \oplus \mathrm{Hom}(K, K \otimes \mathbb{C}^2_{34}) \\
&\quad \to (\mathrm{End}(K) \otimes \mathbb{C}^3) \oplus \mathrm{Lie}U(K) \oplus (\mathrm{Hom}(N,K) \otimes \mathbb{C}^2) \to 0.
\end{aligned}
\tag{5.54}
$$

One once again proves that the degree zero cohomology of this complex is trivial. Since $B_3 = B_4 = 0$ on all solutions, degree one cohomology is the tangent space to instanton moduli space.

The degree two cohomology of this complex forms a real vector bundle of rank $4kn$ over the instanton moduli space. This is known as the obstruction bundle. It turns out in this case that the obstruction bundle is isomorphic to the cotangent bundle of the instanton moduli space–this is a consequence of the extended supersymmetry of $\mathcal{N}=4$ super-Yang-Mills (see [57] for more on this). The virtual tangent space is then defined (in equivariant K-theory) to be the formal difference $H^1 - H^2$ of the nontrivial cohomologies of the above complex, and is denoted T^{vir}.

An equivariant version of this complex is (in hopefully obvious notation)

$$\begin{aligned}
&0 \to \mathrm{Lie}U(K) \to (\sum_{\alpha=1}^{4} q_\alpha^{-1})\mathrm{End}(K) \oplus \mathrm{Hom}(N,K) \oplus q_1^{-1}q_2^{-1}\mathrm{Hom}(K,N) \\
&\to q_1^{-1}q_2^{-1}\mathrm{End}(K) \oplus q_1^{-1}q_3^{-1}\mathrm{End}(K) \oplus q_1^{-1}q_4^{-1}\mathrm{End}(K) \oplus \mathrm{Lie}U(K) \oplus q_3^{-1}\mathrm{Hom}(N,K) \oplus q_4^{-1}\mathrm{Hom}(N,K) \to 0.
\end{aligned} \tag{5.55}$$

The virtual character of this complex may be computed, but requires some care. It is not suitable for localization formulas due to arbitrary choices we have made in assigning signs to the weights of the equations, since they are non-holomorphic. Geometrically, it reflects the fact that the obstruction bundle is a real vector bundle and requires a choice of orientation[1]. See also [41], [42].

What is well-defined and free of choices is the *real* virtual character. If T is the virtual character of the above complex, this is given by $T+T^*$–since it includes both signs on every weight, it is independent of the choice of orientation. The instanton measure can then be defined by $\sqrt{\mathbb{E}[T+T^*]}$. The choice of sign in the square root should be handled with care, and is related to the choice of orientation mentioned above, but we suppress these details and merely present a result which is consistent.

With these preliminaries, we may compute the real virtual character. In what follows *c.c.* denotes

[1] I am grateful to Nikita Nekrasov for a clarifying explanation of this point.

"complex conjugate", or dual character–if one restricts to the compact subgroup of the complex torus acting it is naturally the complex conjugate (note the two copies of Lie$U(K)$ fit together to make $\mathrm{End}(K) = K \otimes K^*$):

$$
\begin{aligned}
&(q_1+q_2+q_3+q_1q_2q_3)KK^* + c.c. + NK^* + c.c. + q_1q_2KN^* + c.c. \\
&\quad - q_1q_2KK^* - c.c. - q_1q_3KK^* - q_2q_3KK^* - c.c. - q_3NK^* - c.c. - q_1q_2q_3KN^* - KK^* - c.c. \\
&=(1-q_3)(NK^* + q_1q_2KN^* - (1-q_1)(1-q_2)KK^*) + c.c. \\
&=(1-q_3)T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n) + c.c.
\end{aligned}
\tag{5.56}
$$

In the last line we have assumed we are considering the tangent space to a fixed point $\underline{\lambda}$. The natural candidate for the holomorphic square root is then $(1-q_3)T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n)$. If one writes w_j, $j = 1,...,2kn$ for the weights in the holomorphic tangent to space $\overline{\mathscr{M}}(k,n)$, the instanton partition function is

$$
Z = \sum_{k\geq 0} \mathbf{q}^k \sum_{\substack{\underline{\lambda} \\ |\underline{\lambda}|=k}} \frac{\prod_{j=1}^{2nk}(\varepsilon_3 + w_j)}{\prod_{j=1}^{2nk} w_j}. \tag{5.57}
$$

If one identifies $\varepsilon_3 = -m$, then this is the partition function for the theory with a hypermultiplet of mass m in the adjoint representation. The reason for this is that the instanton measure can be interpreted as the localization computation of (if m is chosen as the weight in the fiber)

$$
\sum_{\substack{\underline{\lambda} \\ |\underline{\lambda}|=k}} \frac{\prod_{j=1}^{2nk}(m - w_j)}{\prod_{j=1}^{2nk} w_j} = \int_{\overline{\mathscr{M}}(k,n)} e_{\mathbb{T}^{n+3}}(T^*\overline{\mathscr{M}}(k,n)). \tag{5.58}
$$

This is the k-instanton contribution to the partition function of a theory with matter in the adjoint representation (see for example [39]). Note that the Dirac zero modes for a hypermultiplet in the adjoint representation can be identified with the cotangent bundle of instanton moduli space itself.

When $m = 0$, the instanton partition function becomes that of $\mathscr{N} = 4$ super-Yang-Mills theory in the Ω-background. One has (the sum over λ, no underline, is just the sum over all Young diagrams)

$$Z_{\mathscr{N}=4}(\mathfrak{q}) = \sum_{k \geq 0} \mathfrak{q}^k \sum_{\substack{\underline{\lambda} \\ |\underline{\lambda}|=k}} 1 = \Big(\sum_{\lambda} \mathfrak{q}^{|\lambda|} \Big)^n = \Big(\sum_{k=0}^{\infty} \mathfrak{q}^k p(k) \Big)^n = \frac{\mathfrak{q}^{n/24}}{\eta(\tau)^n}. \tag{5.59}$$

$p(k)$ is the number of partitions of the integer k. In the final equality, we have used a standard result for the generating function of partitions of integers and recalled that $\mathfrak{q} = e^{2\pi i \tau}$, where τ is the complexified gauge theory coupling constant. $\eta(\tau)$ is the Dedekind eta function. It is well-known to transform nicely under $\tau \mapsto -1/\tau$:

$$\eta(-1/\tau) = \sqrt{-i\tau}\,\eta(\tau). \tag{5.60}$$

This is the famous S-duality of $\mathscr{N} = 4$ supersymmetric Yang-Mills theory [32], [57].

This construction of generalized ADHM data can be used to compute the partition function of an arbitrary ADE quiver gauge theory. The reason is that the equations for $(B_1, B_2, B_3, B_4, I, J)$ have a symmetry group $SU(2) \times U(1) \times SU(2)$, where (B_1, B_2) transform as a doublet under the first $SU(2)$ and (B_3, B_4) transform as a doublet under the second $SU(2)$. Choosing a discrete subgroup $\Gamma \subset SU(2)$ of the group acting on the $B_{3,4}$ matrices and implementing the orbifold projection on the ADHM data produces the instanton partition function for the $\mathscr{N} = 2$ supersymmetric quiver gauge theory of the same ADE type as Γ. Physically, this is a result of a construction of Douglas and Moore [12] that identifies the low-energy effective theory on the worldvolume of D-branes located at the tip of a $\mathbb{C}^2/\Gamma$ orbifold singularity as a quiver gauge theory of the same ADE type as Γ.

An illustrative case is $\Gamma = \mathbb{Z}_3$, corresponding to the $\mathscr{N} = 2$ quiver gauge theory of type $\widehat{A}_2$. At the level of ADHM data, the vector spaces K and N decompose with respect to the $\mathbb{Z}_3$ action–K splits into K_0, K_1, K_2 and N splits into N_0 and N_1, N_2. The subscripts label the irreducible representation of $\mathbb{Z}_3$ that these spaces carry. The lowercase versions of these letters will be used for the dimensions of the vector spaces. The $B_{1,2}$ matrices split into three copies of themselves which map $K_i \to K_i$ for $i = 0,1,2$. Likewise, the I, J matrices split into maps $I_i : N_i \to K_i$, $J_i : K_i \to N_i$. The $B_{3,4}$ matrices map the different K spaces into one another, specifically B_3 sends $K_i \to K_{i+1}$ and B_4 maps $K_i \to K_{i-1}$.

The full $\mathscr{N} = 4$ theory in the Ω-background[2] has instanton measure determined by the virtual character $(1-q_3)T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n)$. To obtain the instanton measure for the quiver gauge theory, one wishes to take the $\mathbb{Z}_3$-invariant part of this. The term $T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n)^{\mathbb{Z}_3}$ just splits as the sum of the three vector multiplet contributions for gauge groups $U(n_i)$, $i = 0,1,2$ (in the setup of the previous chapter, where N is chosen to be a multiple of the regular representation, $n_0 = n_1 = n_2 = n$, but we continue to use the subscripts to distinguish the vertices). This is the same as a single vector multiplet for the gauge group $U(n_0) \times U(n_1) \times U(n_2)$. In this theory there are three instanton counting parameters, one for each $U(n_i)$ factor of the gauge group, called $\mathfrak{q}_0, \mathfrak{q}_1, \mathfrak{q}_2$.

q_3 carries the defining representation of $\mathbb{Z}_3$, due to the transformation rule for the matrices $B_{3,4}$. This term becomes (the index i is understood $\mod 3$):

$$(q_3 T_{\underline{\lambda}}\overline{\mathscr{M}}(k,n))^{\mathbb{Z}_3} = q_3 \sum_{i=0}^{2} [N_i K_{i+1}^* + q_1 q_2 K_i N_{i+1}^* - (1-q_1)(1-q_2) K_i K_{i+1}^*]. \tag{5.61}$$

This represents the contribution of a matter field coupled to both gauge groups $U(n_i) \times U(n_{i+1})$. To understand this more, we decouple two of them, say $U(n_0)$ and $U(n_2)$ by sending their couplings to zero. This kills all the instanton contributions, so that one sets $K_0 = K_2 = 0$. The contribution

[2] With the mass of the adjoint hypermultiplet turned on, the appropriate name for it is the $\mathscr{N} = 2^*$ theory since the Ω-background in the 34 directions breaks the $\mathscr{N} = 4$ supersymmetry to $\mathscr{N} = 2$.

which remains is (for simplicity we also take the massless limit back to $\mathcal{N}=4$ so that $q_3=1$):

$$N_0 K_1^* + q_1 q_2 K_1 N_2^*. \tag{5.62}$$

This is exactly the contribution of n_0 hypermultiplets in the antifundamental representation of $U(n_1)$, and n_2 hypermultiplets in the fundamental representation. The mass parameters are identified with the Coulomb moduli for the frozen gauge groups. The flavor symmetry can be thought of as the global part of a non-dynamical gauge group. The extra factor $q_1 q_2$ can be absorbed by shifting the mass parameter by an amount $\varepsilon_1+\varepsilon_2$–the convention for whether or not the fundamental contribution includes $\Lambda^2\mathbb{C}^2$ differs in various places in the literature.

Thus, in the full orbifold theory (without decoupling the other gauge groups), the matter content consists of fields in the fundamental representation for $U(n_i)$ and the antifundamental of $U(n_{i+1})$. This reproduces what was found in the previous chapter using the explicit Lagrangian description of the theories.

CHAPTER 6: CROSSED INSTANTONS AND qq-CHARACTER

With the preliminaries finally in place, in the present chapter we exposit the theory of qq-characters. We begin with their geometric origin from moduli spaces describing intersecting branes in string theory, generalizing the ADHM construction discussed in the previous chapter. We then present the formula for the characters in a number of theories, and explain their relation to Nakajima quiver varieties. We briefly discuss the relation of the qq-characters to quantum integrable systems, specifically the illustrative case of Toda chain and $\mathscr{N}=2$ pure supersymmetric Yang-Mills theory.

qq-characters from Intersecting Branes

In the previous chapter, it was observed that instanton calculus in quiver gauge theories is most conveniently organized via a set of generalized ADHM data $(B_1, B_2, B_3, B_4, I, J)$ describing $D(-1)$-branes moving in the background of a stack of $D3$ branes in string theory (for nontrivial ADE quiver gauge theories, one must additionally consider the geometry transverse to the $D3$ branes to be a $\mathbb{C}^2/\Gamma$ orbifold singularity). The B_a matrices, for $a=1,2,3,4$, describe the coordinates of $\mathbb{C}^4$ as seen by the $D(-1)$ branes. To preserve symmetry between the 12 and 34 directions, it is tempting to introduce another stack of $D3$ branes along the 34 directions, which intersect the original stack transversely at the origin of $\mathbb{C}^4$. If the number of $D3$ branes along the 34 directions is w, then one augments the generalized ADHM data with yet another vector space $W \cong \mathbb{C}^w$, and two more maps

$\tilde{I} : W \to K$, $\tilde{J} : K \to W$. Then one imposes the following set of equations on these matrices:

$$\begin{aligned}
[B_1, B_2] + IJ + ([B_3, B_4] + \tilde{I}\tilde{J})^\dagger &= 0 \\
[B_1, B_4] + [B_2, B_3]^\dagger &= 0 \\
[B_1, B_3] + [B_4, B_2]^\dagger &= 0 \\
B_3 I + B_4^\dagger J^\dagger &= 0 \\
B_4 I - B_3^\dagger J^\dagger &= 0 \\
B_1 \tilde{I} + B_2^\dagger \tilde{J}^\dagger &= 0 \\
B_2 \tilde{I} - B_1^\dagger \tilde{J}^\dagger &= 0 \\
\tilde{J} I - \tilde{I}^\dagger J^\dagger &= 0 \\
\left[B_1, B_1^\dagger\right] + \left[B_2, B_2^\dagger\right] + \left[B_3, B_3^\dagger\right] + \left[B_4, B_4^\dagger\right] + II^\dagger + \tilde{I}\tilde{I}^\dagger - J^\dagger J - \tilde{J}^\dagger \tilde{J} &= \zeta \mathbb{1}_K.
\end{aligned} \tag{6.1}$$

The solutions are considered modulo the action of $U(K)$. It is useful for later purposes to relabel the vector space K and K', and write its dimension correspondingly as k'. Then the space of solutions to the above equations is denoted $\overline{\mathscr{M}}(k', n, w)$, and it is called the moduli space of crossed instantons [41], “crossed" referring to the configuration of $D3$ branes. By examining these equations one sees that the virtual or expected dimension of the space is $-2nw$, a negative number. It will be important later that this is independent of k'.

A similar argument as in the case of the generalized ADHM construction for the $\mathscr{N} = 4$ theory yields the following holomorphic description of the space of crossed instantons:

$$\begin{aligned}
[B_1,B_3]=[B_1,B_4]=[B_2,B_3]=[B_2,B_4]&=0\\
[B_1,B_2]+IJ=[B_3,B_4]+\tilde{I}\tilde{J}&=0\\
B_3I=B_4I&=0\\
JB_3=JB_4&=0\\
B_1\tilde{I}=B_2\tilde{I}&=0\\
\tilde{J}B_1=\tilde{J}B_2&=0\\
\mathbb{C}[B_1,B_2]I(N)+\mathbb{C}[B_3,B_4]\tilde{I}(W)&=K'
\end{aligned} \tag{6.2}$$

modulo $GL(K')$. The space $\overline{\mathscr{M}}(k',n,w)$ is in general a complex variety which is singular and consists of various components of distinct dimensions. In this way, its smoothness properties resemble the moduli space of stable maps [25]. The group $GL(N)\times GL(W)\times(\mathbb{C}^\times)^3$ acts on this space, and one can work equivariantly with respect to this action to define integration over this space, by taking the equations above to define a particular virtual fundamental cycle against which one can integrate using equivariant localization. The most succinct definition of the qq character is then the equivariant integral of 1, in the sense defined in the previous chapter:

$$\langle \mathscr{X}_w\rangle=\sum_{k'\geq 0}\mathfrak{q}^{k'}\int_{[\overline{\mathscr{M}}(k',n,w)]^{\mathrm{vir}}}1. \tag{6.3}$$

This is referred to as a qq-character of type $\widehat{A}_0$. Since no orbifold is performed along the 34 directions, this computes a quantity relevant to the $\mathcal{N}=2^*$ theory. Since the second stack of branes transversely interesects the original stack of branes, by integrating out the degrees of freedom coming from the second stack one obtains a local observable in the gauge theory on the stack of $D3$ branes spanning the 12 directions, which is why the above quantity is written as an expectation value of the qq-character observable. What follows will be dedicated to unraveling this definition to obtain more useful information, by integrating out the second stack of branes directly. Performing an orbifold along the 34 directions produces the *qq*-characters relevant for quiver gauge theories.

A special $\mathbb{C}^\times$ action

The group $GL(N)\times GL(W)$ acts on the moduli space of crossed instantons, in particular its subgroup $\mathbb{C}^\times\times\mathbb{C}^\times\subset GL(N)\times GL(W)$ acts on this space. It is easy to see that the diagonal $\mathbb{C}^\times$ group acts trivially on the moduli space, while the anti-diagonal $\mathbb{C}^\times$ acts nontrivially and reflects the coupling of the instantons on each worldvolume to the other. To make better sense of the definition of the *qq*-character, it is useful to first localize the integral (6.3) with respect to this $\mathbb{C}^\times$ action, and then analyze the rest of the symmetries of the problem. The variable $x\in\mathbb{C}$ is introduced as the equivariant parameter for this $\mathbb{C}^\times$ action.

The fixed points of the anti-diagonal $\mathbb{C}^\times$ action have the following structure, as is readily verified. The vector space K' splits as a direct sum, $K'=K\oplus V$, and one can choose K to be a trivial representation of $\mathbb{C}^\times$, and V to carry the defining one-dimensional representation (with multiplicity $\dim V$), so that the elements of V are scaled uniformly by e^x. Likewise, N transforms trivially and W is scaled by e^x under the $\mathbb{C}^\times$ action. The quotient is taken with respect to the subgroup of $GL(K')$ preserving the decomposition $K'=K\oplus V$, which is $GL(K)\times GL(V)$. Then the holomorphic crossed instanton equations (6.2), with this decompositon of K' and quotiented by $GL(K)\times GL(V)$, imply that the matrices (B_1,B_2,I,J) define a $U(n)$ instanton of charge $k=\dim K$, and the matrices $(B_3,B_4,\tilde{I},\tilde{J})$ define a $U(w)$ instanton of charge $v=\dim V$. The fixed point set of this $\mathbb{C}^\times$ action sits inside of the crossed instanton moduli space as

$$\overline{\mathscr{M}}(k',n,w)\supset\bigsqcup_{k'=k+v}\overline{\mathscr{M}}(k,n)\times\overline{\mathscr{M}}(v,w). \tag{6.4}$$

Localization allows the integral (6.3) to be reduced to a sum of integrals over products of two instanton moduli spaces, and the qq-character observable itself $\mathscr{X}(x)$ is defined by integrating the class 1 along the fiber of the projection to $\overline{\mathscr{M}}(k,n)$, in the equivariant sense. This is the precise mathematical statement corresponding to "integrating out the other stack of branes". Note in particular that the qq-character depends on the complex parameter x, and will eventually depend on the rest of the equivariant parameters associated to the maximal torus of symmetries acting on the moduli space $\overline{\mathscr{M}}(v,w)$.

Partial Localization

We now carry out this localization explicitly. The result is an expression for the qq-character of type $\widehat{A}_0$. First, the Atiyah-Bott theorem yields

$$\int_{[\overline{\mathscr{M}}(k',n,w)]^{\mathrm{vir}}} 1 = \sum_{k'=k+v} \int_{\overline{\mathscr{M}}(k,n)\times\overline{\mathscr{M}}(v,w)} \frac{1}{e_{\mathbb{T}}(N^{\mathrm{vir}})} \tag{6.5}$$

Here, $\mathbb{T}$ denotes the maximal torus of symmetries, and $N^{\mathrm{vir}} \in K_{\mathbb{T}}(\overline{\mathscr{M}}(k,n)\times\overline{\mathscr{M}}(v,w))$ is a class in the equivariant K-theory of the fixed locus known as the virtual normal bundle. It is defined by the relation (once again in equivariant K-theory):

$$T^{\mathrm{vir}} = T\overline{\mathscr{M}}(k,n) + T\overline{\mathscr{M}}(v,w) + N^{\mathrm{vir}} \tag{6.6}$$

where T^{vir} is the virtual tangent space, defined as the alternating sum of the cohomologies of the deformation complex associated to the equations (6.1). Once again, because the equations are non-holomorphic and carry a nontrivial obstruction bundle, the obstruction bundle requires a choice of orientation which leads to difficulties with signs, which we again suppress. By a computation identical to those of the previous chapter, one computes, for the real virtual tangent space,

$$\begin{aligned}T_{\mathbb{R}}^{\text{vir}} =&(q_1^{-1}+q_2^{-1}+q_3^{-1}+q_4^{-1})K'K'^{*}+K'N^{*}+q_1^{-1}q_2^{-1}NK'^{*}+K'W^{*}\\ &+q_3^{-1}q_4^{-1}WK'^{*}-(q_1^{-1}q_2^{-1}+q_1^{-1}q_3^{-1}+q_1^{-1}q_4^{-1})K'K'^{*}-(q_3^{-1}+q_4^{-1})K'N^{*}\\ &-(q_1^{-1}+q_2^{-1})K'W^{*}-q_1^{-1}q_2^{-1}NW^{*}-K'K'^{*}+c.c.\end{aligned} \tag{6.7}$$

Inserting $K'=K+e^{x}V$ to extract N^{vir}, one finds after some calculation (and fixing an orientation for the obstruction bundle)

$$T^{\text{vir}}=(1-q_3)T\overline{\mathscr{M}}(k,n)+(1-q_2)T\overline{\mathscr{M}}(v,w)-q_1q_2e^{x}S^{*}\mathscr{C} \tag{6.8}$$

where we have defined

$$\begin{aligned}T\overline{\mathscr{M}}(k,n)&=NK^{*}+q_1q_2KN^{*}-(1-q_1)(1-q_2)KK^{*}\\ T\overline{\mathscr{M}}(v,w)&=WV^{*}+q_3q_4VW^{*}-(1-q_3)(1-q_4)VV^{*}\\ S&=(1-q_1)(1-q_2)K-N\\ \mathscr{C}&=(1-q_3)(1-q_4)V-W.\end{aligned} \tag{6.9}$$

$\mathscr{C}$ is the equivariant K-theory class associated to the tautological complex on $\overline{\mathscr{M}}(v,w)$. From the above, the virtual normal bundle is calculated as

$$N^{\text{vir}}=-q_3T\overline{\mathscr{M}}(k,n)-q_2T\overline{\mathscr{M}}(v,w)-q_1q_2e^{x}S^{*}\mathscr{C}. \tag{6.10}$$

This allows one to write

$$\langle\mathscr{X}_w(x)\rangle=\sum_{k'\geq 0}\mathfrak{q}^{k'}\int_{[\overline{\mathscr{M}}(k',n,w)]^{\text{vir}}}1=\sum_{k+v\geq 0}\mathfrak{q}^{k+v}\int_{\overline{\mathscr{M}}(k,n)\times\overline{\mathscr{M}}(v,w)}c_{\varepsilon_3}(T\overline{\mathscr{M}}(k,n))c_{\varepsilon_2}(T\overline{\mathscr{M}}(v,w))c_{x+\varepsilon}(S^{*}\mathscr{C}). \tag{6.11}$$

We have defined $\varepsilon := \varepsilon_1 + \varepsilon_2$. By noting that the Chern polynomial of $T\overline{\mathscr{M}}(k,n)$ defines the instanton measure in the theory, pulling out this factor one arrives at

$$\mathscr{X}_w(x) = \sum_{v \geq 0} \mathfrak{q}^v \int_{\overline{\mathscr{M}}(v,w)} c_{\varepsilon_2}(T\overline{\mathscr{M}}(v,w)) c_{x+\varepsilon}(S^*\mathscr{C}). \tag{6.12}$$

Note that this has reduced the somewhat formal equivariant integral against $[\overline{\mathscr{M}}(k',n,w)]^{\text{vir}}$ to a tractable object. Note also that S^* is the K-theory class defining the Y-observable upon taking the Chern polynomial, so this object can be interpreted as a certain Laurent series in $Y(x)$, and defines a special x-dependent equivariant cohomology class of the instanton moduli space. It can then be integrated over the space $\overline{\mathscr{M}}(v,w)$ by a further localization computation, an issue to which we return later.

Compactness Theorem

The main result of [41] is that as the parameter $x \in \mathbb{C}$ is varied, the fixed point set in the moduli space of crossed instantons of the torus action remains compact. This implies that the expectation value $\langle \mathscr{X}_w(x) \rangle$ has no poles as a function of the parameter x. This is, at last, the counterpart of the loop equation in the matrix model. $\mathscr{X}_w(x)$ is a combination of Y-observables which has no poles as a function of x, which allows one to fix its analytic behavior by considering limits as $x \to \infty$. This is the source of essentially all interesting applications of the qq-characters, as will be explained shortly. We turn presently to the generalization of this construction to arbitrary quiver gauge theories.

Because of the ability to construct quiver gauge theories by considering D-branes at a singular locus $\mathbb{C}^2/\Gamma$, the qq-characters for the quiver theories (of type ADE) can be obtained via orbifold projection of the moduli space of crossed instantons, by restricting to the fixed locus of the Γ action. The virtual tangent space T^{vir} must be replaced by its Γ-invariant piece. For the term $(1-q_3)T\overline{\mathscr{M}}(k,n)$, this simply defines the instanton measure of the associated quiver gauge theory.

The Γ-fixed part of the moduli space $\overline{\mathscr{M}}(v,w)$ describes the moduli space of instantons on an ALE space of the same ADE type as Γ, as explained in [26], [12]. This space may be given a quiver description, as a so-called Nakajima quiver variety [34]. Denote the quiver variety by $\mathbb{M}_\gamma(\mathbf{v},\mathbf{w})$, where v and w “fractionalize" into vectors $\mathbf{v}$, $\mathbf{w}$, with Vert_γ components where γ is the graph underlying the quiver, which is the affine Dynkin graph of corresponding ADE type. Then the qq-character in the corresponding ADE quiver gauge theory is

$$\mathscr{X}_{\mathbf{w}}(x) = \sum_{\mathbf{v}} \prod_{i\in\mathrm{Vert}_\gamma} \mathfrak{q}_i^{v_i} \int_{\mathbb{M}_\gamma(\mathbf{v},\mathbf{w})} c_{\varepsilon_2}(T\mathbb{M}_\gamma(\mathbf{v},\mathbf{w}))c_{x+\varepsilon}\left(\bigoplus_{i\in\mathrm{Vert}_\gamma} S_i^*\mathscr{C}_i\right). \tag{6.13}$$

S_i and $\mathscr{C}_i$ denote the pieces of S and $\mathscr{C}$ valued in each irreducible representation $i\in\mathrm{Vert}_\gamma$ which correspond to vertices of γ via the McKay correspondence, but they admit a definition purely in terms of the underlying quiver γ and its path algebra. With this in mind, it becomes clear that this formula does not depend on the fact that γ is an affine Dynkin graph of type ADE, so it can be generalized to arbitrary quivers. It is Nekrasov’s eq. 194 in [40], in the case where there are only bifundamental hypermultiplets and no fundamental hypermultiplets at nodes (these can be engineered if desired by starting with bifundamental hypermultiplets and decoupling some gauge group nodes, so that the decoupled nodes become flavor symmetries).

We now turn to the explicit computation of qq-characters in examples. We consider the case of

qq-characters for arbitrary w in quiver gauge theories of type A_1 and the fundamental qq-characters (a name to be explained below) of type A_r. The explicit formulas will illustrate the nomenclature attached to these observables. The qq-characters are usually infinite sums in Y-observables that define Laurent series, but for simplicity in the examples we consider they will always reduce to finite sums.

qq-characters of Type A_1

To obtain the A_1 quiver variety, we can take the $\mathbb{Z}_3$ orbifold projection of $\overline{\mathscr{M}}(v',w')$ and set two of the three V-spaces to zero. The resulting space is the space of pairs $(\tilde{I},\tilde{J}) \in \mathrm{Hom}(W,V) \oplus \mathrm{Hom}(V,W)$ such that $\tilde{I}\tilde{J} = 0$, subject to the stability condition that $\tilde{I}$ is surjective. This space is isomorphic to $T^*Gr(v,w)$, and thus is nonempty only for $0 \leq v \leq w$.

Denote the equivariant parameters for the $GL(W)$ action in the vector space W by $(\nu_1,\ldots,\nu_w)$, and write $\nu := \mathrm{diag}(\nu_1,\ldots,\nu_w)$. Since $\tilde{J}$ also scales under $(\mathbb{C}^\times)^3$ as $\tilde{J} \to q_3^{-1}q_4^{-1}\tilde{J} = q_1q_2\tilde{J}$, the fixed points of the torus action on this space are characterized by the existence of a $g \in GL(V)$ such that

$$\begin{aligned} g\tilde{I}e^{-\nu} &= \tilde{I} \\ q_1q_2e^{\nu}\tilde{J}g^{-1} &= \tilde{J}. \end{aligned} \tag{6.14}$$

These equations imply that $\tilde{J} = 0$, and that V admits a decomposition $V = \oplus_{j\in J}V_j$, where each V_j has eigenvalue ν_j under ν, and $J \subset \{1,2,\ldots,w\}$ is a subset of cardinality v. It is convenient to introduce the notation $[w]$ for the set $\{1,2,\ldots,w\}$. Thus the fixed points of the torus action on this space are labeled by subsets of size v of a set of size w. If we write $\mathbb{M}(v,w) = T^*Gr(v,w)$ for the quiver variety, the character of the tangent space at a fixed point may be easily computed as

$$T\mathbb{M}(v,w) = WV^* + q_1^{-1}q_2^{-1}VW^* - VV^* - q_1^{-1}q_2^{-1}VV^* = \sum_{\substack{i\in I\\ j\in J}} (e^{\nu_i-\nu_j} + q_1^{-1}q_2^{-1}e^{\nu_j-\nu_i}). \quad (6.15)$$

In this sum, I denotes the complementary subset to J, such that $I \sqcup J = [w]$.

Likewise, for the tautological complex in this case one finds

$$\mathscr{C} = V + q_1^{-1}q_2^{-1}V - W = \sum_{j\in J} q_1^{-1}q_2^{-1}e^{\nu_j} - \sum_{i\in I} e^{\nu_i}. \quad (6.16)$$

Taking into account carefully the contributions of the decoupled nodes to $\oplus_{i\in\mathrm{Vert}_\gamma} S_i^*\mathscr{C}_i$, and shifting the mass parameters appropriately, one has in the limit $\widehat{A}_2 \to A_1$,

$$\bigoplus_{i\in\mathrm{Vert}_\gamma} S_i^*\mathscr{C}_i \to S^*\mathscr{C} + q_1^{-1}q_2^{-1}VM^* \quad (6.17)$$

where M is the multiplicity space, which decomposes with respect to the maximal torus of flavor symmetry into eigenspaces with eigenvalues e^{m_i}, $i = 1,\ldots,2n$ (in other words, one has $2n$ fundamental hypermultiplets), and $\mathscr{C}$ (no subscript) is as above.

With these preliminaries in place, one can easily compute the following integral by localization:

$$\begin{aligned} &\int_{\mathbb{M}(v,w)} c_{\varepsilon_2}(T\mathbb{M}(v,w))c_{x+\varepsilon}(S^*\mathscr{C} + e^{-\varepsilon}VM^*) \\ &= \sum_{\substack{J\subset[w]\\ |J|=v}} \prod_{\substack{i\in I\\ j\in J}} \frac{(\nu_i-\nu_j+\varepsilon_2)(\nu_j-\nu_i-\varepsilon_1)}{(\nu_i-\nu_j)(\nu_j-\nu_i-\varepsilon)} \prod_{i\in I} Y(x+\varepsilon+\nu_i) \prod_{j\in J} \frac{P(x+\nu_j)}{Y(x+\nu_j)}. \end{aligned} \quad (6.18)$$

In this equation, $P(x) = \prod_{i=1}^{2n}(x-m_i)$ is a polynomial encoding the mass parameters. Introducing the function

$$S(x) = \frac{(x+\varepsilon_1)(x+\varepsilon_2)}{x(x+\varepsilon)} \tag{6.19}$$

the qq-character of type A_1 may be written

$$\mathscr{X}_w(x) = \sum_{I \sqcup J = [w]} \mathfrak{q}^{|J|} \prod_{i \in I} Y(x+\varepsilon+\nu_i) \prod_{j \in J} \frac{P(x+\nu_j)}{Y(x+\nu_j)} \prod_{\substack{i \in I \\ j \in J}} S(\nu_i - \nu_j). \tag{6.20}$$

In the case $w = 1$, one obtains the so-called fundamental qq-character of type A_1:

$$\mathscr{X}_1(x) = Y(x+\varepsilon) + \mathfrak{q}\frac{P(x)}{Y(x)} \tag{6.21}$$

and the compactness theorem mentioned above implies $\langle \mathscr{X}_1(x) \rangle$ has no singularities in x. Combined with the asymptotics of $Y(x)$ for $x \to \infty$, one concludes this expectation value is a degree n polynomial in x. In the limit of vanishing Ω-background parameters, this encodes the Seiberg-Witten curve of $\mathscr{N} = 2$ supersymmetric Yang-Mills theory with gauge group $U(n)$ coupled to $2n$ hypermultiplets in the fundamental representation (that is, the A_1 quiver gauge theory). To recover the pure Yang-Mills theory, one can decouple the hypermultiplets by sending their mass to infinity, while sending $\mathfrak{q} \to 0$ keeping fixed the quantity $\Lambda^{2n} = \mathfrak{q} \prod_{i=1}^{2n}(-m_i)$. In this way, one reproduces the Seiberg-Witten curve for pure Yang-Mills obtained in the previous section.

Note that the qq-character (6.20) provides an answer to the question posed in the introduction, as it may be expressed in terms of the geometry of the quiver variety via the integration formula (6.13). The qq-characters resemble characters of representations of $\mathfrak{sl}_2$, but are deformed by the various parameters ε, m, ν.

Let us generalize from the case A_1 to the case of A_r quiver theories, but consider only the fundamental qq-characters. These are the characters associated with quiver varieties $\mathbb{M}(\mathbf{v},\mathbf{w})$ of type A_r, where $\mathbf{w}$ has only one nonzero component, which is 1. Note that $\mathbf{v}$, $\mathbf{w}$ are r-component vectors. The quiver variety can be obtained by taking a $\mathbb{Z}_{r+2}$ orbifold projection of $\overline{\mathscr{M}}(v',1)$ and decoupling the vertices associated to the 0 and $r+1$ representation of $\mathbb{Z}_{r+2}$. In fact, these quiver varieties turn out to be merely points, but nonetheless the qq-characters are nontrivial. V decomposes into vector spaces V_i, $i=1,\dots,r$ labeled by the nodes of the A_r quiver, and there is a one-dimensional vector space $W_\ell \cong \mathbb{C}$ at the ℓ-th node, for some $1 \leq \ell \leq r$; the corresponding dimension vector is denoted $\mathbf{w}_\ell$.

Consider the torus fixed points on $\overline{\mathscr{M}}(v',1)$, that is to say, Young diagrams. Because the matrices $B_{3,4}$ map the V_i spaces into one another, it is easy to see that the fixed points on the quiver varieties $\mathbb{M}(\mathbf{v},\mathbf{w}_\ell)$ correspond to partitions of length at most ℓ, that is, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell \geq 0$, with $\lambda_1 \leq r+1-\ell$. The map from the partition λ_i to the sequence of numbers a_i given by

$$a_i = \lambda_i - i + \ell \tag{6.22}$$

maps the set of partitions of length $\leq \ell$ and $\lambda_1 \leq r+1-\ell$ to sequences $(a_i)_{i=1,\dots,\ell}$ of integers such that $r \geq a_1 > a_2 > \cdots > a_\ell \geq 0$. These can be equivalently identified with subsets $I \subset \{0,1,\dots,r\}$ with $|I| = \ell$.

Because of the $\mathbb{Z}_{r+2}$ orbifold procedure, the objects S and $\mathscr{C}$ split as S_i, $\mathscr{C}_i$, for $i=1,\dots r$. One can write $Y_0(x) = P_1(x)$ and $Y_{r+1}(x) = P_r(x)$ for the polynomials $P_{1,r}(x)$ encoding the masses of the fundamental hypermultiplets attached to the first and r-th nodes. The other S_i produce the observables $Y_i(x)$ associated with each $U(n)$ gauge group attached to a node of the A_r Dynkin diagram. One has the formula

$$\mathscr{C}_i = V_i - q_3 V_{i-1} - q_4 V_{i+1} + q_3 q_4 V_i - \delta_{i\ell} W_\ell. \tag{6.23}$$

It is convenient to change notation and use the index $\alpha \in \mathrm{Vert}_\gamma$ to label the objects S_α and $\mathscr{C}_\alpha$ associated to the vertices. Then one has, upon setting $q_4 = q_1^{-1} q_2^{-1}$, $q_3 = 1$ (which can always be assumed up to a shift of some hypermultiplet mass):

$$\bigoplus_{\alpha \in \mathrm{Vert}_\gamma} S_\alpha^* \mathscr{C}_\alpha = \bigoplus_{\alpha \in \mathrm{Vert}_\gamma} -S_\alpha^* (V_{\alpha-1} - V_\alpha) + q_1^{-1} q_2^{-1} S_\alpha^* (V_\alpha - V_{\alpha+1}) - S_\ell^* W_\ell. \tag{6.24}$$

With this choice of q_3, q_4 parameters, introducing the shorthand $q_{12} := q_1 q_2$, one has

$$V_\alpha = \sum_{\substack{(i,j)\in\lambda \\ \ell+j-i=\alpha}} q_{12}^{1-j}. \tag{6.25}$$

V_α is expressed as a sum over boxes in the Young diagram, and in the above equation, $\ell + j - i = \alpha$ is understood as holding mod $r+2$. One has the following elementary combinatorial identity:

$$V_\alpha - V_{\alpha+1} = \sum_{i=1}^{\ell} (q_{12}^{1-i} \delta_{a_i,\alpha} - q_{12}^{1-i} \delta_{\ell-i,\alpha}). \tag{6.26}$$

Substituting this into the above equation, one finds

$$\begin{aligned}\bigoplus_{\alpha \in \mathrm{Vert}_\gamma} S_\alpha^* \mathscr{C}_\alpha &= \bigoplus_{i=1}^{\ell} (q_{12}^{1-i} S_{\ell+1-i}^* - q_{12}^{1-i} S_{a_i+1}^* + q_{12}^{-i} S_{a_i}^* - q_{12}^{-i} S_{\ell-i}^*) - S_\ell^* \\ &= \bigoplus_{i=1}^{\ell} (q_{12}^{-i} S_{a_i}^* - q_{12}^{1-i} S_{a_i+1}^*) - q_{12}^{-\ell} S_0^*\end{aligned} \tag{6.27}$$

This expression is desirable because it clarifies the way the expression on the left hand side decomposes into classes of the Y-observables, which enables one to find the qq-character as an explicit

Laurent series in Y's.

What remains to determine is the class of the tangent space $T\mathbb{M}(\mathbf{v},\mathbf{w}_\ell)$, which can be obtained as the $\mathbb{Z}_{r+2}$-invariant part of the tangent space $T\overline{\mathscr{M}}(v',1)$. From the combinatorial formula (5.30), one has, at the fixed point corresponding to the Young diagram λ

$$T\overline{\mathscr{M}}(v',1) = \sum_{\square\in\lambda}(q_4^{\ell(\square)+1}q_3^{-A(\square)} + q_4^{-\ell(\square)}q_3^{A(\square)+1}). \tag{6.28}$$

Taking the $\mathbb{Z}_{r+2}$-invariant piece means that the only boxes which may contribute are those with hook length $h(\square) = 0 \mod r+2$. However, because the partitions λ are of length $\leq \ell$ and have $\lambda_1 \leq r+1-\ell$, the hook length of every box satisfies the bounds $1 \leq h(\square) \leq r$, so no boxes contribute. From this, one concludes that the quiver variety $\mathbb{M}(\mathbf{v},\mathbf{w}_\ell)$ must be of dimension zero. Since it is known that Nakajima quiver varieties are smooth and connected, it follows that $\mathbb{M}(\mathbf{v},\mathbf{w}_\ell) = \{pt\}$.

Since the quiver varieties are just single points, the integral in (6.13) becomes rather trivial and the main objective is to compute the virtual Chern polynomial of $\oplus_{\alpha\in\mathrm{Vert}_\gamma}S^*_\alpha\mathscr{C}_\alpha$. If the fixed points are indexed by the sequences $r \geq a_1 > a_2 > \cdots > a_\ell \geq 0$, then one has for the qq-character

$$\mathscr{X}_{\mathbf{w}_\ell}(x) = Y_0(x+(1-\ell)\varepsilon)\sum_{(a_i)}\prod_{\alpha\in\mathrm{Vert}_\gamma}\mathfrak{q}_\alpha^{v_\alpha}\prod_{i=1}^{\ell}\frac{Y_{a_i+1}(x+(2-i)\varepsilon)}{Y_{a_i}(x+(1-i)\varepsilon)}. \tag{6.29}$$

Now one observes that

$$\prod_{\alpha\in\mathrm{Vert}_\gamma}\mathfrak{q}_\alpha^{v_\alpha} = (\mathfrak{q}_{a_1}\mathfrak{q}_{a_1-1}\dots\mathfrak{q}_\ell)(\mathfrak{q}_{a_2}\dots\mathfrak{q}_{\ell-1})\dots(\mathfrak{q}_{a_\ell}\dots\mathfrak{q}_1) = \prod_{i=1}^{\ell}(\mathfrak{q}_1\mathfrak{q}_2\dots\mathfrak{q}_{a_i})\times\frac{1}{\prod_{i=1}^{\ell-1}\mathfrak{q}_i^{\ell-i}}. \tag{6.30}$$

It is thus natural to introduce $r+1$ (redundant) variables z_i, $i = 0,1,\dots,r$ via

$$z_i = z_0\mathfrak{q}_1\mathfrak{q}_2\dots\mathfrak{q}_i. \tag{6.31}$$

In terms of these variables, the qq-character becomes

$$\mathscr{X}_{\mathbf{w}_\ell}(x) = \frac{Y_0(x+(1-\ell)\varepsilon)}{z_0 z_1 \dots z_{\ell-1}} \sum_{(a_i)} \prod_{i=1}^{\ell} z_{a_i} \frac{Y_{a_i+1}(x+(1-i)\varepsilon+\varepsilon)}{Y_{a_i}(x+(1-i)\varepsilon)}. \tag{6.32}$$

To complete the simplification of the expression, define the functions (for $i = 0, \dots, r$)

$$\Lambda_i(x) = z_i \frac{Y_{i+1}(x+\varepsilon)}{Y_i(x)} \tag{6.33}$$

and for a finite set I, define the height function $h_I : I \to \mathbb{Z}$ by $h_I(i) = \#\{j | j < i\}$, so that in particular for $I = \{a_1, a_2, \dots, a_\ell\}$, $h_I(a_i) = \ell - i$. Then one has

$$\mathscr{X}_{\mathbf{w}_\ell}(x) = \frac{Y_0(x+(1-\ell)\varepsilon)}{z_0 z_1 \dots z_{\ell-1}} \sum_{\substack{I \subset \{0,1,\dots,r\} \\ |I|=\ell}} \prod_{i \in I} \Lambda_i(x+(h_I(i)-\ell+1)\varepsilon). \tag{6.34}$$

This is exactly the formula presented in [40], see eq. 160-161. It is a deformed version of an elementary symmetric polynomial in the Λ_i variables, which is the same thing as a character of the representation $\wedge^\ell \mathbb{C}^{r+1}$ of $SL(r+1, \mathbb{C})$.

First Application: Seiberg-Witten Geometry of Quiver Gauge Theories

Equipped with the fundamental qq-characters of type A_r, it is simple to characterize the Seiberg-Witten geometry of these gauge theories. Let $\mathscr{X}_\ell(x)$ denote the ℓ-th fundamental qq-character of type A_r. The compactness theorem implies that $\langle \mathscr{X}_\ell(x) \rangle$ is nonsingular in x, and by considering the growth at large x one concludes that it is a polynomial. Introduce the expectation values $y_i = \langle Y_i(x) \rangle$ of the Y-observables, where $i = 1, \dots, r$. In the limit $\varepsilon_{1,2} \to 0$, the expecation values of qq-characters become Laurent series in y_i variables, and the nonperturbative Dyson-Schwinger equations become

$$\chi_\ell(x, y_i) = T_\ell(x) \tag{6.35}$$

for $1 \leq \ell \leq r$, for some polynomials $T_\ell(x)$. This is a set of r independent algebraic equations on the $r+1$ variables $(x, y_1, \ldots, y_r)$, and therefore defines an algebraic curve. This is the Seiberg-Witten curve of the quiver gauge theory, in the description of Nekrasov and Pestun [46]. This is in fact a general phenomenon: for a quiver gauge theory of type ADE, finite or affine, the Seiberg-Witten geometry is determined by the fundamental characters of the corresponding ADE group, in the same manner as above. The qq-characters provide a general context for this phenomenon, as they are constructed at arbitrary values of the Ω-deformation parameters, while the ordinary characters and Seiberg-Witten geometry emerge only in the flat space limit. See the paper [46] for a very detailed analysis of the geometry and physics of these Seiberg-Witten curves.

Application: Wavefunctions of Quantum Integrable Systems

It is also possible to use the theory of qq-characters to make contact with quantum integrable systems. In [48], the authors showed that if one considers $\mathscr{N} = 2$ supersymmetric gauge theories in the special Ω-background where ε_1 is fixed but $\varepsilon_2 \to 0$, then the Seiberg-Witten geometry becomes quantized, with the parameter ε_1 playing the role of Planck's constant. Using the well-known link between Seiberg-Witten theory and integrable systems [21], this provides a mechanism for the quantization of integrable systems using instanton calculus in four-dimensional gauge theory. By introducing a defect supported along a surface into the gauge theory, one is able to produce a wavefunction of the quantum integrable system. The qq-characters in the presence of these defect operators allow one to concretely realize these statements. For this section, consider pure $\mathscr{N} = 2$ super-Yang-Mills with gauge group $U(n)$ for simplicity.

One way to generate a surface defect in gauge theory is to place the theory on an orbifold [44].

For $z = (z_1, z_2) \in \mathbb{C}^2$, consider a surface defect placed along $z_2 = 0$, so that the worldvolume of the defect is the z_1 plane. We consider placing the theory on a $\mathbb{Z}_p$ orbifold for some p: we identify $z_2 \sim \omega z_2$ for ω a primitive p-th root of unity.

On the orbifold, which has fundamental group $\mathbb{Z}_p$, there is now room for an instanton gauge field to approach a nontrivial flat connection at infinity, so there is a decomposition

$$N = \bigoplus_j N_j \otimes R_j \tag{6.36}$$

where R_j, $j = 0, ..., p-1$ are the irreducible representations of $\mathbb{Z}_p$, and N_j is a multiplicity space. The index j is considered mod p. In the representation R_j, the generator ω is sent to ω^j. Likewise, there is a decomposition

$$K = \bigoplus_j K_j \otimes R_j. \tag{6.37}$$

The choice of flat connection (decomposition of N) specifies the surface defect. The case of interest to is when $p = n$ and N is the regular representation, so that $\dim N_j = 1$ for each j. For this reason, the surface defect is known as the regular defect. The partition function in the presence of the defect is denoted by Ψ. Due to the above decomposition of the vector space K, the instanton counting parameter $\mathfrak{q}$ also fractionalizes into parametrs $\mathfrak{q}_i$, which are interpreted as moduli of the surface defect, and on which the surface defect partition function depends.

Since $Y = \mathbb{E}[e^x S^*]$, the decomposition $S = \oplus_j (S_j \otimes R_j)$ allows one to introduce $Y_j(x) = \mathbb{E}[e^x S_j^*]$– thus, the Y observable fractionalizes in the presence of the defect. The original Y-observable is given by $Y(x) = \prod_j Y_j(x)$. It is possible to go back through the derivation of the qq-character observable in the presence of a defect, and one finds for the pure $\mathscr{N} = 2$ theory that the expression

$$\left\langle Y_{\ell+1}(x+\varepsilon) + \mathfrak{q}_\ell \frac{1}{Y_\ell(x)} \right\rangle_{\mathscr{S}} \tag{6.38}$$

has no poles for any $x \in \mathbb{C}$, where the index ℓ runs from 1 to n and is considered modulo n. The subscript signifies that we take the expectation value in the background of a defect; in other words using the orbifolded instanton measure.

The change of variables $\mathfrak{q}_\ell = \Lambda^2 e^{x_\ell - x_{\ell-1}}$ is convenient. Taking the coefficient of x^{-1} of the equation above says that the quantity

$$\varepsilon_1(\sigma_\ell - \sigma_{\ell+1} + \varepsilon_2 k_\ell) + \frac{\varepsilon_1^2}{2}[(k_\ell - k_{\ell+1})^2 + k_\ell - k_{\ell+1}] - a_{\ell+1}\varepsilon_1(k_\ell - k_{\ell+1}) + \mathfrak{q}_\ell \tag{6.39}$$

has vanishing expectation value. The variables k_ℓ are the instanton charges associated with each representation of $\mathbb{Z}_n$, in other words the instanton contributions are weighted as $\prod_\ell \mathfrak{q}_\ell^{k_\ell}$. σ_ℓ is a complicated object, the details of which we suppress. Summing over ℓ (recalling that it is considered modulo n) one has

$$\varepsilon_1\varepsilon_2 \sum_\ell \langle k_\ell \rangle_{\mathscr{S}} + \frac{\varepsilon_1^2}{2}\sum_\ell \langle (k_\ell - k_{\ell+1})^2 \rangle_{\mathscr{S}} - \varepsilon_1 \sum_\ell a_{\ell+1}\langle (k_\ell - k_{\ell+1}) \rangle_{\mathscr{S}} + \sum_\ell \mathfrak{q}_\ell \langle 1 \rangle_{\mathscr{S}} = 0. \tag{6.40}$$

The surface defect partition function is a function $\Psi(\Lambda, x_1, \ldots, x_n, a_1, \ldots, a_n, \varepsilon_1, \varepsilon_2)$. In terms of these variables one has, as is easily verified by considering the orbifolded instanton measure,

$$\begin{aligned} \Big\langle \sum_\ell k_\ell \Big\rangle_{\mathscr{S}} &= \frac{1}{2}\Lambda \frac{\partial \Psi}{\partial \Lambda} \\ \langle (k_\ell - k_{\ell+1}) \rangle_{\mathscr{S}} &= \frac{\partial \Psi}{\partial x_\ell} \end{aligned} \tag{6.41}$$

so that one is left with

$$\left(\frac{1}{2}\varepsilon_1\varepsilon_2 \Lambda \frac{\partial}{\partial \Lambda} + \frac{\varepsilon_1^2}{2}\sum_\ell \left(\frac{\partial}{\partial x_\ell} - \frac{a_{\ell+1}}{\varepsilon_1} \right)^2 - \frac{1}{2}\sum_\ell a_\ell^2 + \Lambda^2 \sum_\ell e^{x_\ell - x_{\ell-1}} \right) \Psi = 0. \tag{6.42}$$

Now we take the limit $\varepsilon_2 \to 0$ and substitute in the expected asymptotics

$$\Psi \sim \prod_\ell e^{\frac{a_{\ell+1}}{\varepsilon_1} x_\ell} e^{\frac{1}{\varepsilon_2}\widetilde{W}} \left(\psi + \mathscr{O}(\varepsilon_2)\right) \tag{6.43}$$

to find

$$\left(\frac{1}{2}\varepsilon_1 \Lambda \frac{\partial \widetilde{W}}{\partial \Lambda} + \frac{\varepsilon_1^2}{2}\sum_\ell \frac{\partial^2}{\partial x_\ell^2} - \frac{1}{2}\sum_\ell a_\ell^2 + \Lambda^2 \sum_\ell e^{x_\ell - x_{\ell-1}}\right)\psi = 0. \tag{6.44}$$

Rearranging, this says that

$$\left(\frac{\varepsilon_1^2}{2}\sum_\ell \frac{\partial^2}{\partial x_\ell^2} + \Lambda^2 \sum_\ell e^{x_\ell - x_{\ell-1}}\right)\psi = E\psi \tag{6.45}$$

where

$$E = \frac{1}{2}\sum_\ell a_\ell^2 - \frac{\varepsilon_1}{2}\Lambda \frac{\partial \widetilde{W}}{\partial \Lambda}. \tag{6.46}$$

This establishes that the function ψ is a wavefunction of periodic Toda chain. The quantity $\widetilde{W}$ is the effective twisted superpotential introduced in [48], and it determines the exact energy levels of the quantized integrable system. These manipulations illustrate the utility of qq-characters, as the qq-characters in the presence of surface defects are able to allow one to construct wavefunctions of nontrivial quantum integrable systems.

The applications of the qq-characters that we have presented only scratch the surface of the theory. Readers are directed to [43], [49] for more, in particular in connection with the so-called BPS/CFT correspondence relating the BPS sector of $\mathscr{N} = 2$ gauge theories to two-dimensional conformal field theories.

CHAPTER 7: CONCLUSION

In this thesis we have reviewed some of the modern developments in instanton calculus in $\mathcal{N} = 2$ supersymmetric gauge theories. We surveyed the background material, both the results from matrix models which helped to inspire these developments as well as the basics of modern instanton calculus, including the ADHM construction, partial compactification of instanton moduli spaces, and the Ω-deformation. We then explained, using motivation from string theory, the moduli space of crossed instantons and provided first-principles calculations of the qq-character observables in gauge theory. These techniques allowed us to give short and direct derivations of the Seiberg-Witten geometry of all theories under consideration, in a manner that paralleled the large N limit of matrix models. We also briefly sketched a relation to the theory of quantum integrable systems.

It would be interesting to attempt to extend the philosophy of nonperturbative Dyson-Schwinger equations beyond the realm of supersymmetric theories, as a way to study symmetries of more interesting quantum field theories such as pure Yang-Mills. However, this is likely to be extraordinarily challenging. It would also be interesting to consider in more detail the Yangian symmetries which underlie the qq-characters; as explained in an early chapter, these symmetries can arise in geometric representation theory and one has an explicit description of the action of the algebra on the cohomologies of quiver varieties. While the gauge theory calculations hint at their presence via the qq-characters, there is (at least as far as the author knows) no direct construction of the action of these algebras on the quantum field theory in question. It could be rewarding to attempt to pursue this direct construction. Nekrasov's localization technique can also be generalized beyond the realm of four dimensional instantons, to consider counting supersymmetric configurations in various theories of various dimensions (the localization calculations in Donaldson-Thomas theory [30], [31] can be interpreted as a six-dimensional version of instanton counting, for example). Perhaps there is an analog of qq-characters in all of these "BPS state counting" theories that organizes non-obvious symmetries hiding behind the partition functions.

Bibliography

[1] M. F. Atiyah and R. Bott. "The moment map and equivariant cohomology". In: *Topology* 23.1 (1984), pp. 1–28.

[2] M. F. Atiyah et al. "Construction of instantons". In: *Physics Letters A* 65.3 (1978), pp. 185–187.

[3] N. Berline and M. Vergne. "Classes caractéristiques équivariantes. Formule de localisation en cohomologie équivariante". In: *C. R. Acad. Sci. Paris Sér. I Math.* 295 (1982), pp. 539–541.

[4] R. Bott. "Homogeneous Vector Bundles". In: *Annals of Mathematics* 66.2 (1957), pp. 203–248.

[5] S. Cordes, G. Moore, and S. Ramgoolam. "Lectures on 2D Yang-Mills theory, equivariant cohomology and topological field theories". In: *Nuclear Physics B - Proceedings Supplements* 41.1 (1995), pp. 184–244.

[6] E. Corrigan and P. Goddard. "Construction of instanton and monopole solutions and reciprocity". In: *Annals of Physics* 154.1 (1984), pp. 253–279.

[7] P. Deligne and D. S. Freed. *Supersolutions*. 1999. DOI: 10.48550/ARXIV.HEP-TH/9901094. URL: https://arxiv.org/abs/hep-th/9901094.

[8] R. Dijkgraaf and C. Vafa. "Matrix models, topological strings, and supersymmetric gauge theories". In: *Nuclear Physics B* 644.1 (2002), pp. 3–20.

[9] S. K. Donaldson. "An application of gauge theory to four-dimensional topology". In: *Journal of Differential Geometry* 18.2 (1983), pp. 279–315.

[10] S. K. Donaldson. "Instantons and geometric invariant theory". In: *Communications in Mathematical Physics* 93.4 (1984), pp. 453–460.

[11] M. Douglas. *Branes within Branes*. 1995. arXiv: `hep-th/9512077 [hep-th]`.

[12] M. Douglas and G. Moore. *D-branes, Quivers, and ALE Instantons*. 1996. arXiv: `hep-th/9603167 [hep-th]`.

[13] J. Duistermaat and G. Heckman. "On the Variation in the Cohomology of the Symplectic Form of the Reduced Phase Space." In: *Inventiones mathematicae* 69 (1982), pp. 259–268.

[14] B. Eynard. *Counting Surfaces*. Progress in Mathematical Physics. Springer Basel, 2016.

[15] B. Eynard, T. Kimura, and S. Ribault. *Random matrices*. 2018. arXiv: `1510.04430 [math-ph]`.

[16] B. Eynard and N. Orantin. "Invariants of algebraic curves and topological expansion". In: *Communications in Number Theory and Physics* 1.2 (2007), pp. 347–452.

[17] W. Fulton and J. Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York, 1991.

[18] D. Gaiotto. "$\mathscr{N} = 2$ Dualities". In: *Journal of High Energy Physics* 2012.8 (2012).

[19] D. Gaiotto, G. W. Moore, and A. Neitzke. "Wall-crossing, Hitchin systems, and the WKB approximation". In: *Advances in Mathematics* 234 (2013), pp. 239–403.

[20] D. Gieseker. "On the Moduli of Vector Bundles on an Algebraic Surface". In: *Annals of Mathematics* 106.1 (1977), pp. 45–60.

[21] A. Gorsky et al. "Integrability and Seiberg-Witten exact solution". In: *Physics Letters B* 355.3-4 (1995), pp. 466–474.

[22] P. Griffiths and J. Harris. *Principles of algebraic geometry*. Wiley classics library. New York, NY: Wiley, 1994.

[23] K. Hori et al. *Mirror symmetry*. Vol. 1. American Mathematical Soc., 2003.

[24] E. Kiritsis. *String theory in a nutshell*. Vol. 21. Princeton University Press, 2019.

[25] M. Kontsevich. “Enumeration of Rational Curves Via Torus Actions”. In: *The Moduli Space of Curves*. Ed. by R. H. Dijkgraaf, C. F. Faber, and G. B. M. van der Geer. Birkhäuser Boston, 1995, pp. 335–368.

[26] P. Kronheimer and H. Nakajima. “Yang-Mills instantons on ALE gravitational instantons”. In: *Mathematische Annalen* 288.1 (1990), pp. 263–307.

[27] B.F Logan and L.A Shepp. “A variational problem for random Young tableaux”. In: *Advances in Mathematics* 26.2 (1977), pp. 206–222.

[28] A. Losev, N. Nekrasov, and S. Shatashvili. “The Freckled Instantons”. In: *The Many Faces of the Superworld: Yuri Golfand Memorial Volume*. Ed. by M. Shifman. World Scientific, 2000, pp. 453–475.

[29] M. Mariño. *Chern-Simons Theory, Matrix Models, and Topological Strings*. International series of monographs on physics. Clarendon Press, 2005.

[30] D. Maulik et al. “Gromov-Witten theory and Donaldson-Thomas theory, I”. In: *Compositio Mathematica* 142.5 (2006), pp. 1263–1285.

[31] D. Maulik et al. “Gromov-Witten theory and Donaldson-Thomas theory, II”. In: *Compositio Mathematica* 142.5 (2006), pp. 1286–1304.

[32] C. Montonen and D. Olive. “Magnetic monopoles as gauge particles?” In: *Physics Letters B* 72.1 (1977), pp. 117–120.

[33] V. P. Nair. *Elements of Geometric Quantization and Applications to Fields and Fluids*. 2016. arXiv: `1606.06407 [hep-th]`.

[34] H. Nakajima. “Instantons on ALE spaces, quiver varieties, and Kac-Moody algebras”. In: *Duke Mathematical Journal* 76.2 (1994), pp. 365–416.

[35] H. Nakajima. *Lectures on Hilbert Schemes of Points on Surfaces*. University lecture series. American Mathematical Society, 1999.

[36] H. Nakajima. *Introduction to quiver varieties – for ring and representation theoriests*. 2016. DOI: 10.48550/ARXIV.1611.10000. URL: https://arxiv.org/abs/1611.10000.

[37] Hiraku Nakajima. *Quiver varieties and finite dimensional representations of quantum affine algebras*. 1999. DOI: 10.48550/ARXIV.MATH/9912158. URL: https://arxiv.org/abs/math/9912158.

[38] N. Nekrasov. "Trieste Lectures on Solitons in Noncommutative Gauge Theories". In: *Superstrings and Related Matters: Proceedings of the Trieste 2000 Spring Workshop On, ICTP, Trieste, Italy, 27 March-4 April 2000*. Ed. by C. Bachas et al. World Scientific, 2001, pp. 141–205.

[39] N. Nekrasov. "Seiberg-Witten Prepotential from Instanton Counting". In: *Advances in Theoretical and Mathematical Physics* 7.5 (2003), pp. 831–864.

[40] N. Nekrasov. "BPS/CFT Correspondence: Non-perturbative Dyson-Schwinger Equations and qq-characters". In: *Journal of High Energy Physics* 2016.3 (2016).

[41] N. Nekrasov. "BPS/CFT Correspondence II: Instantons at Crossroads, Moduli and Compactness Theorem". In: *Advances in Theoretical and Mathematical Physics* 21 (2017), pp. 503–583.

[42] N. Nekrasov. "BPS/CFT Correspondence III: Gauge Origami Partition Function and qq-Characters". In: *Communications in Mathematical Physics* 358.3 (2017).

[43] N. Nekrasov. *BPS/CFT Correspondence V: BPZ and KZ Equations from qq-characters*. 2017. arXiv: 1711.11582 [hep-th].

[44] N. Nekrasov. "BPS/CFT Correspondence IV: Sigma Models and Defects in Gauge Theory". In: *Letters in Mathematical Physics* 109.3 (2018), pp. 579–622.

[45] N. Nekrasov and A. Okounkov. “Seiberg-Witten Theory and Random Partitions”. In: *The Unity of Mathematics: In Honor of the Ninetieth Birthday of I.M. Gelfand*. Ed. by P. Etingof, V. Retakh, and I. M. Singer. Birkhäuser Boston, 2006, pp. 525–596.

[46] N. Nekrasov and V. Pestun. *Seiberg-Witten geometry of four dimensional $\mathscr{N} = 2$ quiver gauge theories*. 2012. arXiv: `1211.2240 [hep-th]`.

[47] N. Nekrasov and A. Schwarz. “Instantons on Noncommutative $\mathbb{R}^4$, and $(2,0)$ Superconformal Six Dimensional Theory”. In: *Communications in Mathematical Physics* 106.1 (1998), pp. 45–60.

[48] N. Nekrasov and S. Shatashvili. “Quantization of Integrable Systems and Four-Dimensional Gauge Theories”. In: *XVIth International Congress on Mathematical Physics* (2010).

[49] N. Nekrasov and A. Tsymbaliuk. *Surface defects in gauge theory and KZ equation*. 2021. arXiv: `2103.12611 [hep-th]`.

[50] V. Pestun. “Review of localization in geometry”. In: *Journal of Physics A: Mathematical and Theoretical* 50.44 (2017).

[51] N. Seiberg and E. Witten. “Electric-magnetic duality, monopole condensation, and confinement in $\mathscr{N} = 2$ supersymmetric Yang-Mills theory”. In: *Nuclear Physics B* 426.1 (1994), pp. 19–52.

[52] J-P. Serre. “Linear representations and Kählerian homogeneous spaces of compact Lie groups”. In: *Bourbaki seminar: years 1951/52 - 1952/53 - 1953/54, exhibitions 50-100*. Bourbaki seminar 2. Talk: 100. Société mathématique de France, 1954.

[53] M. Shifman. *Advanced Topics in Quantum Field Theory: A Lecture Course*. Cambridge University Press, 2012.

[54] J. Teschner. *New Dualities of Supersymmetric Gauge Theories*. Mathematical Physics Studies. Springer International Publishing, 2016.

[55] D. Tong. *TASI Lectures on Solitons*. 2005. arXiv: hep-th/0509216 [hep-th].

[56] L. Tu. *Introductory Lectures on Equivariant Cohomology*. Annals of Mathematics Studies. Princeton University Press, 2020.

[57] C. Vafa and E. Witten. “A strong coupling test of S-duality”. In: *Nuclear Physics B* 431.1-2 (1994), pp. 3–77.

[58] Michela Varagnolo. *Quiver Varieties and Yangians*. 2000. DOI: 10.48550/ARXIV.MATH/0005277. URL: https://arxiv.org/abs/math/0005277.

[59] A. Vershik and S. Kerov. “Asymptotics of the Plancherel measure of the symmetric group and the limit form of Young tableaux”. In: *Soviet Math. Dokl.* 18 (1977), pp. 527–531.

[60] E. Witten. “Topological quantum field theory”. In: *Communications in Mathematical Physics* 117.3 (1988), pp. 353–386.

[61] E. Witten. “Introduction to Cohomological Field Theories”. In: *International Journal of Modern Physics A* 06.16 (1991), pp. 2775–2792.

[62] E. Witten. “Sigma models and the ADHM construction of instantons”. In: *Journal of Geometry and Physics* 15.3 (1995), pp. 215–226.

[63] N.M.J. Woodhouse. *Geometric Quantization*. Oxford mathematical monographs. Clarendon Press, 1997.

[64] A. Zee. *Quantum field theory in a nutshell*. Vol. 7. Princeton university press, 2010.